PERSONNEL SELECTION
IN THE
BRITISH FORCES

PERSONNEL SELECTION
IN THE
BRITISH FORCES

PHILIP E. VERNON
M.A., Ph.D.
and
JOHN B. PARRY
B.A., Ph.D.

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4.

FIRST PRINTED 1949

658.311
VER

AGENTS OVERSEAS

AUSTRALIA, NEW ZEALAND AND SOUTH SEA ISLANDS
W. S. SMART P.O. Box 120, SYDNEY, N.S.W.
Showroom: 558, George Street.

CANADA
CLARKE, IRWIN & CO., LTD.,
103, St. Clair Avenue West, TORONTO, 5.

EGYPT AND SUDAN
DINO JUDAH NAHUM P.O. Box 940, CAIRO.
Showroom: 44, Sharia Sherif Pasha.

FAR EAST
(*including China and Japan*)
DONALD MOORE 22, Orchard Road, SINGAPORE.

INDIA
ORIENT LONGMANS LTD.,
BOMBAY: Nicol Road, Ballard Estate.
CALCUTTA: 17, Chittaranjan Ave.
MADRAS: 36A, Mount Road.

SOUTH AFRICA
H. B. TIMMINS P.O. Box 94, CAPE TOWN.
Showroom: 58–60, Long Street.

Bureau Ednl. [illegible] Research
DAVID HARE TRAINING COLLEGE
Dated 15.3.55
Accs. No 588

Printed & Bound in England for the UNIVERSITY OF LONDON PRESS, LTD.,
by HAZELL, WATSON & VINEY, LTD., Aylesbury and London

PREFACE

DURING the second World War, scientific methods of personnel selection were applied in the Royal Navy, the Army and A.T.S., and the Royal Air Force on a scale quite unprecedented in Britain. This development has naturally aroused widespread interest, and a desire for further information beyond that provided by articles in the popular Press or in technical psychological journals. The present book attempts, therefore, to give an account of the procedures used and the results obtained. It is addressed both to the industrialist or educationist who hopes that the methods may be of value in the peace-time selection of employees, students or pupils, and to the student of psychology or to the personnel officer who wishes to know what psychological discoveries were made and what techniques were found most valuable. If for this reason it falls between two stools, we must apologise; it is not easy to write a simple, readable and interesting description for the former and at the same time to give the latter all the technical details that he needs. Statistical terms, formulæ and calculations have been kept to the barest minimum and will, it is hoped, be published elsewhere.

Our primary aim has been to describe such novel contributions as were made in the Fighting Services. But although space has not allowed a comprehensive survey of peace-time selection and guidance, we have tried to view our methods and results in this broader setting, and to integrate them with recent developments in civilian vocational and educational psychology.

After an historical introduction, Part I is concerned with the organisation of selection, the general procedures employed, and the work of psychologists in the Royal Navy, Army and A.T.S., and the R.A.F. Part II takes up the principles of selection and guidance which have evolved both from pre-war investigations and from war-time experience. It reviews the techniques of interviewing and testing, together with evidence of their merits and defects, in some detail. Each chapter is preceded by a general summary or abstract, and the main conclusions relevant to peace-time work are brought together in the final chapter.

It should be made clear that two important topics are almost wholly neglected. First, we have not been able to cover the work on selection problems by psychologists outside the British Fighting Services during 1939–47. Much was done, for example, by members of the Cambridge University Psychology Department, University College (London), and by other teams and individuals. Again, to attempt to describe the even more extensive developments of personnel selection in the American Forces would have been quite impossible at the time of writing. Secondly, we have not tried to deal with military applications of psychology apart from selection—for example, studies of training methods, of the design and layout of equipment, of the effects of conditions of warfare on morale, and so forth—whether conducted by psychologists inside or outside the Services. It may safely be stated that at least as valuable contributions to the efficiency of the Forces can be, and have been, made by psychologists in such fields as by those concerned with selection and guidance. The same, of course, is true in industry and education.

The reader's attention is drawn to two appendices which list, first the naval, military and air force abbreviations employed, and secondly, the numbers or abbreviated titles of the main psychological tests referred to in the text. Dates are inserted in the text after the authors' names to serve as references to their books or articles, which are listed in the bibliography at the end of the book.

We wish to express our thanks to Mr. Alec Rodger, Col. B. Ungerson, Dr. H. J. Eysenck and Mr. B. S. Morris for reading parts, or the whole, of the manuscript and for their suggestions and criticisms ; and to Mrs. Dorothy Vernon for help in preparing the index. Acknowledgments are also due to the American Psychological Association, Inc., and the National Institute of Industrial Psychology for permission to reprint portions of articles (Vernon, 1947a, 1947b, Parry, 1947) which have appeared in *The American Psychologist* and *Occupational Psychology*.

Finally, it should be stated that this book is published by permission of the Lords Commissioners of the Admiralty, the Army Council and the Air Council, but the responsibility for any statements of fact or opinions rests solely with the authors.

CONTENTS

PREFACE 5

Part I

I THE RISE OF VOCATIONAL PSYCHOLOGY . . . 11
II PERSONNEL SELECTION IN THE ROYAL NAVY . . 25
III OTHER RANK SELECTION IN THE ARMY AND A.T.S. . 39
IV ARMY OFFICER SELECTION 52
V SELECTION IN THE ROYAL AIR FORCE . . . 67

Part II

VI GENERAL PRINCIPLES OF VOCATIONAL CLASSIFICATION . 85
VII THE VALUE OF VOCATIONAL CLASSIFICATION PROCEDURES 103
VIII THE BIOGRAPHICAL QUESTIONNAIRE . . . 131
IX THE INTERVIEW AND ITS VALIDITY . . . 143
X PRINCIPLES OF PSYCHOLOGICAL TESTING . . . 165
XI EFFECTS OF PRACTICE, AGE AND OTHER FACTORS ON TEST SCORES 181
XII GENERAL SURVEY OF THE VALUE OF VOCATIONAL TESTS . 204
XIII VERBAL INTELLIGENCE AND EDUCATIONAL TESTS . 220
XIV NON-VERBAL AND MECHANICAL TESTS . . . 234
XV SPECIAL APTITUDE, AND TEMPERAMENT, TESTS . 247
XVI MAIN R.A.F. SELECTION FINDINGS 261
XVII CONCLUSIONS 281

Appendices

I ABBREVIATIONS EMPLOYED IN THE TEXT OR IN APPENDIX II 293
II THE MAIN PSYCHOLOGICAL TESTS USED IN THE FORCES 295
BIBLIOGRAPHY 302
INDEX 313

LIST OF DIAGRAMS

CHART INDICATING THE MAIN ORIGINS AND THE COURSE OF DEVELOPMENT OF VOCATIONAL PSYCHOLOGY . 13
1. SCATTERGRAMS OF CORRELATION COEFFICIENTS WITH SELECTION RATIOS OF 50 PER CENT. AND 20 PER CENT. 129
2. PERCENTAGE OF WASTAGE REMAINING WITH DIFFERENT SIZES OF CORRELATION 130
3. PREDICTIVE VALUE OF A CLERICAL APTITUDE TEST . 179
4. NINETIETH, FIFTIETH AND TENTH PERCENTILE T2 SCORES OF MEN EMPLOYED IN DIFFERENT BRANCHES OF THE NAVY. 211
5. SPECIMEN ITEM AS IN THE PROGRESSIVE MATRICES TEST 235
6. SQUARES TEST. 237
7. FIGURE CONSTRUCTION TEST 237
8. PAPER FORMBOARD TEST 238
9. MECHANICAL COMPREHENSION TEST ITEMS . . . 241
10. PERCENTAGE OF PILOT TRAINING WASTAGE BEFORE GRADING 263
11. PILOT TRAINING WASTAGE BY CALENDAR QUARTERS . 264
12. FLIGHT TEST (R.A.F.) E.F.T.S. 264
13. FLIGHT TEST (R.C.A.F.) SPECIAL GRADING EXPERIMENT . 265
14. SECOND FLIGHT TEST MARKS, APRIL, 1942. FIRST 100 CADETS GRADED AT EACH SCHOOL 270
15. RELATION OF GRADING PERFORMANCE TO SUBSEQUENT ACCIDENT RATE IN PILOT TRAINING . . . 271
16. RELATION OF GRADING TO AGE DIFFERENCES AND FAILURE RATE 272

PART I

The organisation of selection, the general procedures employed and the work of psychologists in the Royal Navy, Army and A.T.S., and the Royal Air Force.

CHAPTER I

THE RISE OF VOCATIONAL PSYCHOLOGY

Abstract.—Psychology and the other social sciences can help in the solution of many problems of the modern world, both in war and peace. The scientific approach is particularly valuable whenever prediction of human behaviour enters, as in educational and vocational selection. The chart on page 13 indicates the main origins and the course of development of vocational psychology, which are described in this chapter.

It is a truism that the solution of the problems of individuals and of societies in the modern world depends more on progress in the social sciences, including psychology, sociology and economics, than on further advances in the physical and biological sciences, engineering and medicine. The importance of psychology becomes particularly obvious during periods of emergency such as the two World Wars, since the need for making the most effective possible use of human as well as of mechanical resources is then realised. It is the object of this book to describe the developments that occurred in war-time in applying psychology to the classification of British manpower, and especially of mind or brain power.* Scarcely less urgent, however, are the problems of adjusting the individual in peace-time to the educational and industrial systems of the society in which he lives, and of adjusting these systems to the legitimate demands of individuals. Hence we shall attempt to show how far experience gained in the guidance and selection of recruits may be carried over into civilian practice.

Now the social sciences are at the same time the most immature and the most intractable of all the sciences precisely because they are concerned with man. Individuals differ too widely in their tastes, talents and temperaments to be readily amenable to scientific analysis and control. As each one is usually convinced that he knows what is best for himself and his neighbours, he resents being studied, classified and regulated by psychologists and economists. He fears that this involves forfeiting his free choice and being directed in all his doings from birth to the grave by an army of scientific experts—a prospect even more abhorrent than his present subjection to teachers, employers and civil servants. This suspicion however, is hardly warranted, for several reasons.

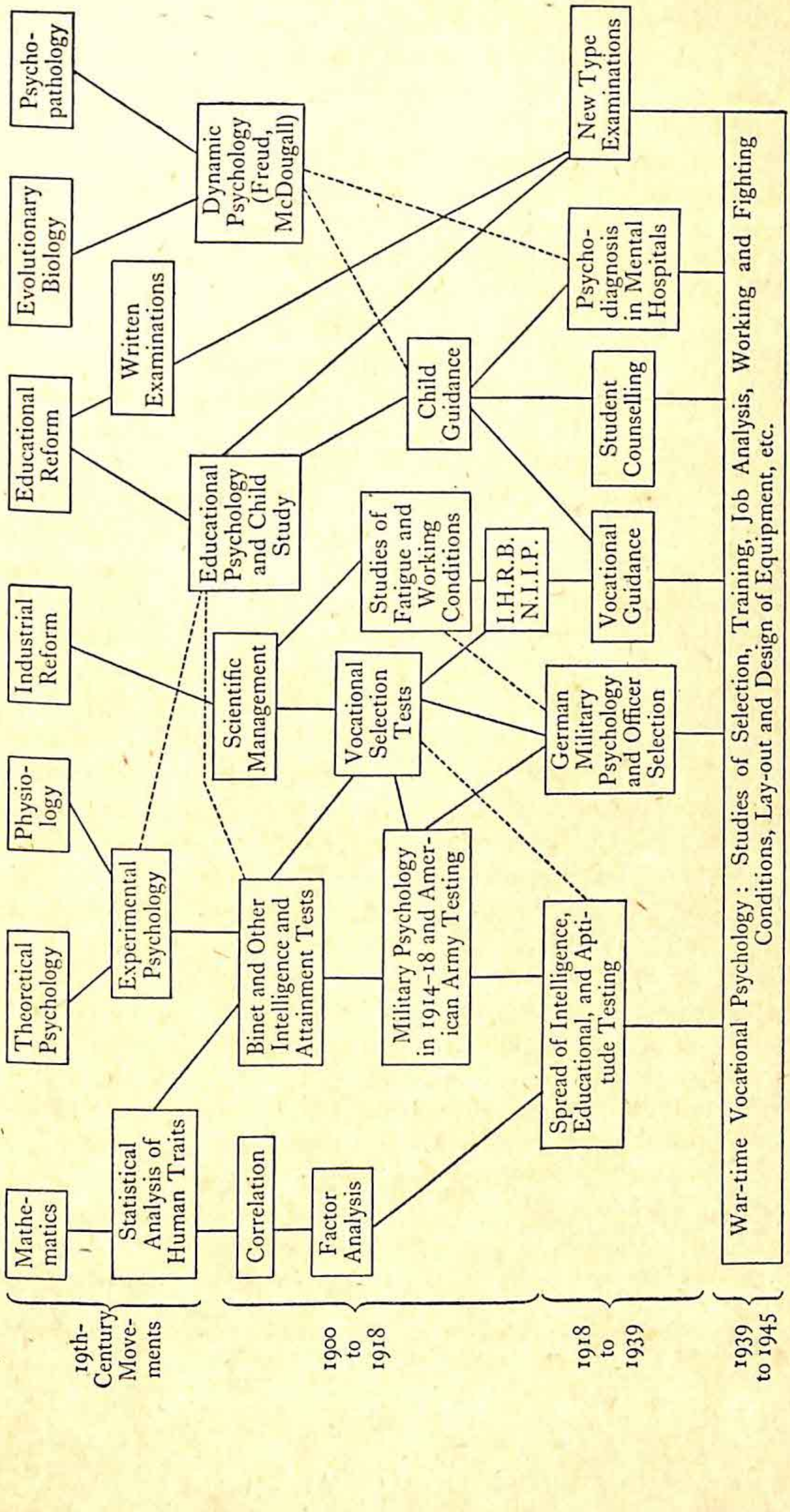
---

* This conception of "human engineering" is borrowed from Yoakum and Yerkes' book on mental tests in the American Army (1920), and Yerkes' memorandum to the U.S. Secretary of War (1941).

First, the psychologist admits, more readily indeed than does the opinionated layman, that he is not omniscient about human nature. He prefers not to pronounce definite decisions except in those rare instances where he has strong scientific proof of their validity. In fact, his caution is often apt to seem rather irritating. Secondly, the good psychologist recognises and allows as much scope as possible for individuality. He never willingly dictates a course of action, but attempts rather to clarify the situation confronting the individual, and to draw attention to all the relevant factors, so that the individual may himself arrive at a wise and satisfying solution. (His method of tackling problems within a group or organisation such as an industrial firm, or the army, is similar. It certainly does not approximate to psychological totalitarianism.)* Thirdly, it must be admitted that man's freedom is already greatly circumscribed. Decisions are made over his head by his parents, teachers and employers, usually on much more arbitrary and fallible grounds than those which the psychologist, with his specialised training and ingenious techniques, takes into account. Expressed in another way, the life-history of an individual is punctuated by a series of predictions. When his mother smacks him, she predicts that this will make him behave better in future; his success or failure in the Special Place examination predicts his suitability or unsuitability for secondary education; similarly, his acceptance or rejection for a job, or in marriage, are essentially prophecies. Psychological investigations have proved that many of these predictions are incorrect, and that they are potent sources of unhappiness, delinquency and neurosis. Predictions by psychologists themselves are certainly not perfect, but they are constantly being checked and improved. In many fields, such as the upbringing of children, in educational and vocational guidance and selection, their superiority to the haphazard and unsystematic predictions

---

* cf. the articles on psychotherapy, industrial and educational investigations by members of the Tavistock Clinic, Jaques, *et al.* (1947).


19th-Century Movements
Mathematics
Theoretical Psychology
Physiology
Industrial Reform
Educational Reform
Evolutionary Biology
Psychopathology
Statistical Analysis of Human Traits
Experimental Psychology
Written Examinations
Dynamic Psychology (Freud, McDougall)
1900 to 1918
Correlation
Factor Analysis
Binet and Other Intelligence and Attainment Tests
Scientific Management
Educational Psychology and Child Study
Vocational Selection Tests
Military Psychology in 1914–18 and American Army Testing
Studies of Fatigue and Working Conditions
Child Guidance
I.H.R.B. N.I.I.P.
1918 to 1939
Spread of Intelligence, Educational, and Aptitude Testing
German Military Psychology and Officer Selection
Vocational Guidance
Student Counselling
Psychodiagnosis in Mental Hospitals
New Type Examinations
1939 to 1945
War-time Vocational Psychology: Studies of Selection, Training, Job Analysis, Working and Fighting Conditions, Lay-out and Design of Equipment, etc.

of everyday life has been demonstrated. In others, such as international relations and economic policies, psychologists are convinced that scientific investigations would help, but are seldom so rash as to claim that their knowledge is sufficiently far advanced to provide immediate remedies for the country's or the world's ills.

Historical Outline

Only a bird's eye view will be attempted here of the development of the branch of applied psychology with which we are concerned. Viteles (1932) provides an excellent history of industrial psychology; a Board of Education Report (1924) describes the rise of mental testing, and Flugel (1933) outlines the wider background. Until halfway through the nineteenth century, psychology was an entirely theoretical subject—a branch of philosophy—and the foundation of the first laboratory at Leipzig by Wundt in 1879 epitomised its conversion into an experimental science of human thought and behaviour. Such work as that of Helmholtz on vision, hearing and reaction time (quickness of response), and Galton's fertile explorations of individual differences in intellect and mental imagery (Inquiries Into Human Faculty, 1883), contributed to this new outlook. Among Wundt's many pupils were the Americans, J. McK. Cattell, who in the 1890's investigated the wide range of differences on simple tests of sensory and motor capacities, and Stanley Hall, who fostered some of the first systematic studies of childhood. Contemporaneously the nineteenth century witnessed a great increase in humanitarian attitudes towards the insane and mentally defective, and in education and in industry, and these stimulated the demand for better methods of diagnosing and predicting human qualities and capacities.

Education

As universal education spread, the tremendous differences in innate learning abilities came to be realised. With the desire to give higher education to those who would profit most instead of merely to a privileged social class, there arose the system of written examinations. The Civil Service adopted examinations to eradicate the evils of nomination in 1870, and the professions similarly reformed their qualifying procedures at about this time. While such examinations marked a great advance over oral tests by inspectors in schools and "disputations" in the universities, later work has revealed their defectiveness as selection techniques owing to unreliability and subjective standards of marking (cf. Vernon, 1940). The first tests of intelligence, namely Ebbinghaus's Completion Test and Binet's famous scale, were constructed at the request of the education authorities of Breslau and Paris respectively at the turn of the century. These tests and their numerous progeny made more accurate the segregation of mentally defective children for special education, also the discrimination between backward children who are innately dull and those whose retardation is due to physical, emotional or other causes. Tests of attainment in spelling, reading and arithmetic, and scales for grading hand-writing and composition, were constructed by Thorndike, Courtis and others in America around 1910. These were standardised so that a child's performance could be compared with the average for his age, and accurate surveys could be made of the relative achievement of classes or schools. Thorndike also established the pattern for scientific studies of teaching methods. The superiority or inferiority of a new device over stock methods could be demonstrated by the progress made among comparable groups of children, instead of by personal impression and arguments based on tradition. In 1913 Burt was appointed by the London County Council as first psychologist to an education authority and initiated his series of researches into intelligence and attainment tests, into delinquency and backwardness, and into statistical methods of analysing psychological data, which formed the basis of so much subsequent British work in applied psychology.

We must here retrace our steps and mention the important contributions to the conception of mental measurement of mathematicians such as Quetelet, who demonstrated the applicability of the normal or Gaussian frequency curve to human characteristics, and Galton and Pearson whose technique of correlation made it possible to measure the extent of agreement between scores on two or more tests or examinations (cf. p. 104). From this evolved factor analysis, which is used for classifying the main types of human abilities. In early writings on psychology and education frequent reference is made to faculties or mental powers such as attention, memory, observation, etc. Spearman's researches from 1904 onwards largely discredited these and demonstrated the importance of "*g*", the general intellectual factor, underlying all our abilities. Burt and Thomson in England and Thurstone and others in America greatly extended factorial techniques and demonstrated the existence of additional factors or types of ability—verbal, numerical, mechanical, spatial, etc.

Educational psychology was profoundly influenced again by nineteenth-century developments in evolutionary biology and in psychopathology. McDougall's writings on the instincts and sentiments and Freud's studies of the unconscious motives underlying human nature, greatly increased our understanding of the difficult or nervous child, the delinquent and the neurotic. The first clinic for children was established by Witmer at the University of Pennsylvania in 1896, but this was mainly concerned with treating scholastic difficulties. In 1909 Healy commenced his studies of behaviour problem cases and delinquents in Chicago. Burt was the chief originator of similar work in England, followed by Boyd and Drever in Scotland.

Industry

While many early psychological studies of fatigue and of conditions affecting learning were relevant to industrial work, the first important experiments in industry itself were those of Taylor and Gilbreth, in America, around 1910. The stress laid by these experts in "scientific management" and their followers on the increased output resulting from more efficient working methods tended to antagonise workers and the Trade Unions. Although industrial psychologists have been at least as interested in the satisfaction of the worker as in his output, this suspicion has unfortunately persisted and was responsible for some lack of co-operation in the Second World War between the Ministry of Labour and psychologists in the Forces. Nevertheless an important series of researches on conditions in British factories was begun by H. M. Vernon, under the Health of Munition Workers Committee, in the first war, and continued by the Industrial Fatigue Research Board (later the Industrial Health Research Board)—a branch of the Medical Research Council. These were concerned with working hours, rest pauses, heating, lighting, accidents, fatigue and boredom, and methods of selection and guidance.

As early as 1905 a French psychologist, Lahy, was studying the abilities required for typewriting and tramway driving, but the realisation of the general principles of vocational psychology is usually attributed to Muensterberg, who worked in Germany till 1911 and then in America. He outlined the functions of psychology in contributing to industrial efficiency and in serving the interests both of employees and employers, and himself developed tests for selecting tramcar drivers, telephone operators and others. Gradually the view made headway that the choice of the right man is no less important than the choice of the right machine. It was estimated that, at a typical job, the most efficient workers are fully three times as competent as the least efficient.

1914-1918

This brings us to the first World War, when psychologists in Germany, Britain, France, Italy and America were engaged on the development of tests for night vision, for the selection of pilots and observers, drivers and telegraphists, for the anti-submarine service, etc., as well as on other contributions to military psychology which Burt (1942) has described. Few of these tests have survived, and in most countries psychologists were too few in number to undertake large-scale schemes or to have much influence on the service authorities. But an outstanding landmark in the history of vocational psychology was the application, during 1917–18, of group intelligence tests (devised by Otis and other American psychologists) to nearly two million American army recruits (cf. Yoakum and Yerkes, 1920, Memoirs of the National Academy of Sciences, 1921). Up to 2,000 a day were tested in any one centre, in batches of one to five hundred at a time. A battery of verbal tests —Army Alpha—was used with the majority, but a non-verbal series—Army Beta—was given to the 30 per cent. who could barely read English, and adaptations of the Binet scale, or performance tests, to borderline defectives. The value of the test scores was clearly demonstrated, and they could be used for :

(1) Selecting men of superior ability for officer or N.C.O. training, or other special duties.
(2) Selecting duller men for labour or special training units, and for eliminating the lowest grades.
(3) Balancing the distribution of ability between different units.
(4) Allocating men into groups which would be homogeneous in ability and thus helping instruction.

Though these applications were permissive, not obligatory, psychologists had less trouble in persuading the Army to adopt them than in dissuading it from attaching over-much importance to test scores and neglecting physical, personality, and other relevant factors. Owing to lack of time relatively little attempt was made to test aptitudes for specialised training, but extensive work was carried out on trade tests for assessing the proficiency of recruits from skilled and semi-skilled civilian trades.

1919-1939

This demonstration of the possibilities of large-scale selection procedure provided an immense stimulus to group-testing in America. Intelligence tests soon came into general use in schools and universities, while achievement and new-type scholastic tests, with similar make-up and items to those of intelligence tests, largely supplanted the unreliable essay-type examination. Organisations such as the American Council on Education through its Co-operative Test Service provided each year new forms of intelligence and achievement tests standardised to yield comparable scores at the college level. In the 1930's various methods of automatic scoring were devised. With the most widely used system (adopted for pencil-and-paper tests by the U.S. Forces in the second World War), testees indicate their responses to each item by drawing a line with a carbon pencil on a special answer sheet. When the sheet is placed in an electrical machine, current passes through correctly drawn lines and yields the total score instantaneously.

In this country intelligence tests have spread widely in primary schools. For example Moray House constructs new tests yearly for education authorities to apply in selection for secondary education at 11+. In 1932 and again in 1947 the complete 11-year-old population of Scotland was tested by the Scottish Council for Research in Education. Individual tests such as revisions of the Binet scale, performance tests and educational tests (e.g. Burt's, 1921) are generally employed in Child Guidance Clinics. But there was relatively little testing in secondary schools or universities, or at the adult level until 1941, and new-type examinations have gained little ground in spite of Ballard's able advocacy (1923), and the exposure by Hartog and Rhodes (1935) and others of the inadequacies of conventional examinations. Perhaps the most notable outcomes of psychological work in 1914–18 were the setting up of the Industrial Fatigue Research Board, and the foundation by Myers of the National Institute of Industrial

Psychology in 1921. The N.I.I.P. is an independent organisation whose chief activities include investigations for particular firms into their problems of selection, training and working conditions, and the giving of vocational guidance to individual applicants—mostly boys and girls leaving school—who desire advice on suitable careers. So far as its limited funds would permit, it too has conducted researches, for example into tests of mechanical aptitude, of motor-driving and other abilities, and into practicable techniques of guidance for school-leavers in large areas. While certain firms such as Rowntrees (cf. Northcott, 1945) have employed psychologists in their personnel departments, the majority of industrial and vocational psychologists in this country have been trained by the I.H.R.B. and N.I.I.P., and a great deal of the work of these bodies between the wars was directly relevant to the problems with which British psychologists were called upon to cope in 1941. Selection tests were applied to apprentices in the Army and the Royal Air Force, for example by Cox (1928), Stanbridge (1936), and Farmer and Chambers (1936). The writings of Burt *et al.* (1926) and Macrae (1932) have been particularly influential in the field of guidance.

The first attempts to give vocational guidance to children are said to have been those of Parsons in America in about 1900, but many school teachers have long accepted the responsibility of advising their leaving pupils unofficially regarding suitable employment in the light of their knowledge of the pupils' abilities and characters. Juvenile Employment Committees, working either under the Ministry of Labour or under Local Education Authorities, are nowadays concerned with vocational placement, and do their best to integrate the views of the school, the child's and parents' interests, and local labour requirements. But much evidence has accumulated as to the untrustworthiness of these haphazard procedures and the superiority of the more scientific guidance methods worked out by the N.I.I.P. (cf. Chapter VII). Mention should be made also of the Appointments Boards or Committees for placing university students in jobs, the first of which was established at Oxford University in 1899. A growing number of secondary schools have part-time careers masters or mistresses, given introductory training at the N.I.I.P., who, in the quite inadequate time usually allotted, try to survey the pupils systematically and to give them guidance on leaving. Such activities are much better developed in American high schools and colleges, where they are known as "Student Counselling". They include advice to students on their work and educational careers, financial, health and family difficulties, and personality maladjustments as well as vocational plans (cf. Williamson, Darley, *et al.*, 1937, 1941). The vast "Advisement and Guidance" programme of the United States Veterans' Administration is on similar lines (cf. Scott and Lindley, 1946). Two to three thousand psychologists are engaged at some 400 centres in studying the potentialities and needs of ex-Service men and helping them to readjust to civilian life and to reach sound decisions regarding jobs or further training, etc. It is desirable to point out the dangers of applying psychology on so large a scale, in view of the difficulties of finding really suitable personnel and of training them adequately.

Educational and vocational guidance have often been given, in this country, by Child Guidance Clinics. During the 1920's clinics spread rapidly, both in America and Britain, with the support of the Commonwealth Fund. Their staff usually comprised a psychiatrist in charge of diagnosis and treatment, a social worker to explore the home background, and a psychologist to survey the children's abilities and to carry out remedial educational work. Play and speech therapists were also often attached. This system has, and still does, work successfully. But just as vocational guidance has generally been conducted by psychologists and teachers, so child guidance is being taken over more and more by psychologists with teaching qualifications—a psychiatrist usually being available for consultation on cases showing serious neurotic tendencies. The establishment of minimum qualifications for Associateship and Fellowship of the British Psychological Society in 1941, and of its Committee of Professional Psychologists (Mental Health) in 1944–5, has helped to ensure that guidance is being given by properly trained persons. Another field in which the relationship and functions of psychologist and medical officer are gradually becoming clarified is that of psychodiagnosis among mental hospital patients or in adult clinics. When psychological departments were first introduced in mental hospitals, it was all too common for patients to be sent by the psychiatrist with a demand for "an I.Q.". With the development of projection and other diagnostic tests (cf. Rapaport, 1944), the psychologist is nowadays better qualified than his medical colleague to explore the particular mental functions in which deterioration may have occurred, and to analyse the general personality structure—provided he has received thorough theoretical, practical, and interne training. Such training is difficult to come by in this country as yet, but in America several hundred clinical psychologists were employed in the army during the war, and at least as many are now working for the Veterans' Administration on diagnosis, psychotherapy and research with neuropsychiatric cases, in addition to those in civilian posts (cf. Miller, 1946; Hutt and Milton, 1947).

A vast amount of research, which we cannot attempt to summarise here, was published between the wars on topics of the utmost concern to vocational psychologists, e.g. the relative influence of heredity and environmental conditions on intelligence and other abilities; differences between the sexes, national and occupational groups, and age changes; new tests and diagnostic techniques, particularly in the field of personality; statistical methods such as applications of analysis of variance to psychological data; studies of the structure, and relations between, abilities and traits; analyses of jobs and of working methods. One matter, however, which we must not neglect is the development of vocational psychology in Germany.

German Military Psychology

The lead given by Muensterberg was taken up, as Viteles (1926–30) has shown, more enthusiastically by German than by British and American industry. Numerous organisations, including Krupps, Zeiss and the Berlin Tramways had psychological departments largely engaged in devising and applying selection tests to applicants for employment. Moreover, while military psychology lapsed completely in English-speaking countries during the 1920's and 30's, Germany built on what had been achieved and actively developed it in the reconstitution of her Forces from 1927 onwards. It has been claimed, probably with justification, that the speed with which the Luftwaffe was re-created was largely due to the excellent quality of the human material selected for it. As Farrago (1941), Ansbacher (1941) and—more briefly—Burt (1942) describe, each Army Corps had its psychological section, staffed by carefully chosen personnel who had undergone military as well as psychological training. Military psychology claimed to embrace personnel

Accessioned No. 588

selection, problems of training and indoctrination, discipline and morale among recruits, design of machinery and equipment, analysis of army employments and of leadership, study of the psychological characteristics of foreign countries and propaganda for home and abroad. But post-war enquiries indicate that very little was actually done in fields other than selection. And even here, although tests were devised for pilots, tank drivers, radio operators, sound detectors and other specialists, the emphasis was less on technical ability than on soldierly qualities—keenness, loyalty, perseverance and adaptability. This tendency reached its acme in the choice of officers.

Four or five candidates at a time underwent a three-day procedure before a board consisting of a colonel (who made the final decision), a medical officer and three psychologists. They were tested on the first and third days, and during the intervening day their behaviour was closely observed. While many of the tests resemble those used in Britain and America, little use was made of any scores; instead, qualitative features such as the manner of performing the tasks were stressed, the whole aim being to build up a realistic picture of the personality as a whole, not a compendium of separate traits and abilities. Nor was any one personality type desired. The examination of the candidates, it was stated, should keep as closely as possible to the concrete requirements of the officer's life, and the psychologists were required to use their practical understanding of human nature as much as their technical skill. This examination fell into four main sections:

(1) Intellectual and other abilities. Ordinary verbal and performance tests were given, usually individually, tests of mechanical comprehension, and a sorting test to reveal the quality of intellectual generalisation or abstraction. A cinema film was shown for candidates to describe, their answers showing their powers of observation and imagination. Practical problems were posed in the form of essays to bring out their planning capacities.

(2) "Action analysis," i.e. character and temperament. A complex choice reaction time test was claimed to show power of sustained attention or distractibility, speed of learning, fatiguability, and emotional control. In the study of personality, a test involving interpretation of pictures (Thematic Apperception) was used, together with a projection test

based on the completion of simple drawings. Other, military, tests involved the drilling of recruits and instructing them in some task or lecturing to them, also carrying out complex orders requiring quickness of uptake, agility and endurance, and improvisation in emergencies. Candidates were encouraged or severely criticised, in order that the effects of these social stimuli might be observed.

(3) Expression analysis. Facial expression was observed, and often photographed by a concealed cine camera, during conversation or tests involving distractions or painful stimuli. Speech, handwriting and literary style were also analysed in detail, and the significance of each mode of expression for the total personality considered.

(4) Life-history. The candidate's party record was studied, and data were collected by questionnaire and interview on his past development, his aims and ambitions, his self-evaluation and social attitudes. At the close there was a discussion of some topic by all the candidates, designed to show up their competitiveness and other social reactions. Detailed reports were drawn up for the consideration of the superior officer.

Although the selectees, both officers and specialists, were followed up at intervals and their progress compared with the psychologists' judgments, this comparison was of a subjective kind. Excellent or good agreement was alleged in 80–90 per cent. of cases, but there was no scientific validation either of the selection procedure as a whole, or of the deductions made from the separate tests. In other words the claim that such and such a test revealed will-power, imagination, or other vague faculties was entirely *a priori*, and in the absence of empirical substantiation, German psychologists had little defence when they fell into disfavour. Actually their prestige was already declining in 1939 and their work was entirely stopped in 1941. It is generally believed that this was due to political prejudice, but Davis (1947) attributes it rather to the inherent weaknesses of their approach.

Conclusion

With the opening of the second World War, British psychologists were eager to put their training and experience to useful service. But the prestige of vocational psychology among naval and military authorities and in Government circles was much lower in this country than in America, and there was no strong central body for co-ordinating their efforts and drawing attention to the contributions they could make. Much valuable work was done during 1939–41 by individuals and by university teams, particularly for the Royal Air Force, also on such civilian psychological problems as evacuation of children, effects of air-raids, and public opinion and morale surveys. The next four chapters describe how the Services came to realise their need for "human engineering" and how this was met.

CHAPTER II

PERSONNEL SELECTION IN THE ROYAL NAVY

Abstract.—Selection in the Royal Navy was somewhat unscientific and unsystematic up till 1941, when the Senior Psychologist's Department of the Admiralty was established. The primary object of the department was to assist in selection of recruits at combined recruiting centres, and in their allocation to suitable branches at entry establishments. The staff of the department, their training, and the tests and other techniques employed, are described. Most of the qualified personnel were civilians, but interviewing, allocation, etc., were largely carried out by Wren petty officers or personnel selection officers. Important features were the provision of information on jobs to recruits, the matching of supply with the demands of all the branches, and re-allocation of failures. Personnel work was later developed in depots, barracks and specialist schools, among Fleet Air Arm mechanics, Marines and artificers; and a scheme of selection was integrated with the training of officer candidates. Follow-up or validation investigations, other research activities, studies of training, documentation, "work-simplification" and opinion surveys, are outlined.

Up till 1941 such selection as existed in the Royal Navy was based mainly on educational examinations and on interview. Scholastic performance played a large part in the admission of cadets and of artificer apprentices, that is of future officers and skilled tradesmen, also a smaller part in the acceptance of continuous service and special service ratings. War-time volunteers and conscripts ("Hostilities Only") ratings were medically examined and interviewed at Combined Recruiting Centres (C.R.C.s). Both their acceptance or rejection and the category or branch to which they were allocated were decided by the naval recruiter, who was usually a pensioner chief petty officer or petty officer. For certain classes of tradesmen such as electrical mechanics trade tests were given. Such specialists as gunnery, torpedo, Radar and Asdic (anti-submarine) ratings were nominally chosen after a period at sea on the basis of a recommendation from their commanding officer, but, owing to shortages of experienced men around 1941, the major proportion were merely put on training courses by the drafting officers at the naval depots or barracks without any enquiry into their suitability. Neither the recruiters, nor the officers concerned with selection, had any training in interviewing, and with the increasing complexity of modern equipment and the multiplication of specialist categories, they could seldom be expected to know much about the jobs for which they were selecting. Many of them employed unstandardised spelling or mental arithmetic questions as tests of education and intelligence*. Several training schools had home-made group intelligence tests; fortunately, perhaps, little attention was paid to their results. Some schools made approaches to trained psychologists for assistance. Thus in December, 1940, one of the writers was asked to devise selection tests for Asdic ratings, because the quality of men sent for training was so poor. His battery (described on p. 231) was applied for the rest of the war, and it certainly prevented a number of potential failures from cluttering up the courses. But its validity was not high enough for it to pick all the potential successes. In all, many hundreds of men made the lengthy journey to training schools in Scotland, only to be rejected without even starting training. Moreover, the procedure took no account of the equally clamant needs of numerous other naval branches. These defects could only be overcome by a centrally organised scheme of selection.

The system which worked adequately in the small peace-time Navy could barely cope with the vast war-time expansion. Although in 1941 the Navy was in the fortunate position of being able to reject four out of every five applicants, numerous complaints were made by entry establishments, depots, and specialist schools about the quality of trainees. For example among seamen torpedomen the failure rate in one of the main schools rose from the pre-war figure of under 10 per cent. to 31 per cent. in 1942, but was brought down to 10 per cent. again by 1944. Not only were men without the necessary mechanical or other aptitudes breaking the hearts of the training staffs, but also many with excellent qualifications were wasted on relatively unskilled work. The effects on

* The classic example was the recruiter who was heard saying to a volunteer: "What? Spell Egypt with a J? You're illegible for the Navy."

the morale of recruits in general were as serious as the shortages of competent personnel.

Partly owing to the urgency of these difficulties, partly through the initiative of the Consultant in Neuropsychiatry, a committee was appointed by the Admiralty in March, 1941, to study the selection and allocation of Hostilities Only recruits. A. Rodger (1945) has described the visits paid by the committee to combined recruiting centres, entry establishments, depots, training schools, and ships of the Fleet; also how its recommendations were accepted in the main, the Senior Psychologist's (S.P.) Department being established under the Second Sea Lord, and new schemes instituted at recruiting centres and entry establishments before the end of the year.

The Staff of the Senior Psychologist's Department

The backbone of the psychological staff consisted of eight civilian psychologists provided by the National Institute of Industrial Psychology. Five other men with psychological qualifications joined later, either as civilians or as R.N.V.R. officers, and two trained psychiatric social workers served throughout most of the war. But the bulk of the personnel had little or no psychological training at the start. At the peak period they numbered about 300, including 50 Wren first, second or third officers (personnel selection officers), 120 Wren petty officers or C.P.O.'s (recruiting assistants), and some 30 Wrens or Marines engaged on clerical duties, 15 sick berth attendants (assistants to neuropsychiatrists), and about 20 civilian clerks and statistical assistants (cf. Expert Committee, 1947).

The Wren recruiting assistants were mostly drawn direct from civil life, where they had been teachers, employment officers, social workers and the like. Though only a small proportion were university graduates, they were nearly all of university calibre and were extremely carefully chosen for the work in view. They received a fortnight's intensive training in their future duties, in which the Director of Naval Recruiting and the Consultants in Neuropsychiatry and Ophthalmology co-operated. They had to acquire skill in interviewing, testing technique and record-keeping, and to obtain some acquaintance with the educational systems of Britain and with the nature of civilian occupations. Their work was regularly supervised by the psychiatric social workers.

From their number the most suitable were later chosen for promotion to personnel selection officer (P.S.O.). Though they had by now gained a good deal of relevant experience, they required much additional training in interviewing and in knowledge of civilian and naval jobs. For it was they who had to decide whether a man claiming to be a fitter, a lathe-setter, a radio mechanic, etc., was properly qualified; or whether a recruit would most suitably adapt to training as electrical mechanic, stoker, telegraphist, seaman, writer, or cook and so on; and their employment recommendations were almost always accepted. Their training consisted mainly in an apprenticeship to an industrial psychologist who was already carrying out actual selection, but included visits to specialist schools, conferences and refresher courses at the Admiralty, often a period in a Government training centre to learn about the handling of bench and machine tools, and spells at the Admiralty to gain acquaintance with the administrative system and recording arrangements.

Certain points deserve comment. First, the Senior Psychologist—the director of the organisation, whose status was similiar to that of the directors of several other Admiralty departments—was himself a psychologist. Moreover, all the policy and administration were decided and carried out by psychologists. Although psychologists as such did not at that time receive professional recognition in the Civil Service, as scientists or doctors did, yet there was a strong scientific atmosphere in the department and an insistence on professional standards, which undoubtedly helped to foster good work. And although psychologists and P.S.O.s had no executive authority in the Royal Navy, but acted throughout as technical advisers, there was far less frustration than is common when psychologists work in Government departments or in the Services under non-professional direction (cf. Chase, 1947).

Secondly, most of the staff, including the fully-qualified psychologists, worked in the field—at entry establishments, specialist schools, depots, etc., and even the few trained personnel who were attached to Admiralty headquarters made frequent excursions in order to keep in touch with the realities of selection. When any new job was undertaken, the guiding principle was for one of the professional psychologists to get it going on the spot and to seek proof of the effectiveness of his methods, then to make these methods as simple and foolproof as possible in order to hand them over to P.S.O.s or other less skilled personnel, so that he might be free to turn his attention to yet another problem (cf. Straker, 1944). The same principle is highly appropriate in peace-time industry and education where, as shown below (p. 100), industrialists and teachers will need to be trained to carry out most selection or other psychological work.

The fact that nearly 90 per cent. of the personnel were women was due primarily to the non-availability of men of similar intellectual calibre. But the success with which they acquired knowledge of male occupations, and handled thousands of individual recruits or tested large groups, and became accepted as useful members of the staffs of almost all-male establishments, was very striking. It is more because of the reduction in the size of the W.R.N.S. than because of any inability of women to carry out selection, that much of the post-war work has been taken over by R.N. and R.N.V.R. personnel. This, too, carries a lesson for peacetime selection and guidance.

The Recruiting Centre Scheme

The function of Wren recruiting assistants was to supply the recruiters with factual information on the basis of which they could more effectively decide the suitability or otherwise of candidates for the Royal Navy. This they did by applying:

(1) A biographical questionnaire, designed to bring out educational and occupational history, leisure interests and experience of leadership.

(2) The Progressive Matrices test, S.P. Test 0.

(3) Selected plates from the American Optical Company's version of the Ishihara and Stilling colour blindness tests.

(4) A short interview, mainly for purposes of clarifying written questionnaire responses, but usually including also a series of informal questions about nervousness, dieting and the like for detecting recruits who ought to be referred for psychiatric examination.

Careful records were kept and weekly returns sent to headquarters. The questionnaires and test results of accepted candidates were forwarded to the recruits' entry establishments. Although no fixed pass-mark was set on the intelligence test, few men from the bottom 10 per cent. of the population were accepted unless, in the opinion of the recruiter, they possessed other exceptional qualifications for the Navy. By 1942–3 the Navy had to take nearer 80 per cent. than 20 per cent. of candidates, but later it became possible to raise the minimum intelligence standards for the majority of recruits to approximately the median for the general population. An analysis of returns and reasons for rejection over the period December, 1941—May, 1942, showed that out of every 100 candidates :

74·4 were accepted
13·4 were rejected on medical grounds
4·7 were rejected because their trade was not required, or because of insufficient trade experience
2·6 were rejected on grounds of low Matrices scores
1·1 were rejected on grounds of low educational level or illiteracy (judged from the questionnaire)
3·8 were rejected on other grounds

In the five years following the institution of this scheme, approximately one million candidates were put through it. But with the demobilisation of recruiting assistants at the end of the war, an interim scheme was introduced, which the recruiters themselves were trained to operate. A battery of six self-administering 10–15 minute tests was provided, together with suggested standards for the various categories to which recruits could be allocated (radio mechanic, writer, seaman, stoker, etc.). Several parallel forms of each of the following tests were constructed:

A. Mechanical information and comprehension.
B. Mathematics. C. Spelling. D. Abstraction.
E. Advanced arithmetic. F. English.

Only tests A and B are given to all conscripts and volunteers, C and D being added with recruits for the Regular Navy, and E and F serving as educational examinations for certain classes of tradesmen.

The Entry Establishment Scheme

In 1941 the majority of H.O. recruits were sent to one of four entry establishments to receive their elementary training in seamanship. In each of these an industrial psychologist was placed to carry out the second phase of selection procedure (cf. Rodger, 1945; Jennings, 1947). Later, almost all recruits went to a single establishment, H.M.S. *Royal Arthur*. They spent a fortnight here before beginning their training, in "kitting up", and in medical and psychological examinations. In addition to one resident and two or more visiting psychologists, there were 15 P.S.O.s to deal with some 1,200 men a week, and 25 Wren ratings who were engaged mainly on test correcting and records. The resident psychologist was either a civilian or a R.N.V.R. officer. Also working with his department were a psychiatrist and a R.N. executive officer, whose respective functions were to interview men referred by P.S.O.s as possible neurotics or as potential officers. There were six main features of the procedure in H.M.S. *Royal Arthur*, namely:

1. *Provision of Information.*—Posters, photographs, and bulletins describing the work, rates of pay, prospects and necessary qualifications, were displayed for incoming recruits to study and discuss. A preliminary talk by the psychologist or a P.S.O. both described the selection procedure, and drew attention to the various naval jobs, emphasising that it was up to every recruit to do the most difficult job of which he was capable, and not merely one which he thought would give him a pleasant time. Films and trade test pieces were also sometimes demonstrated.

2. *More Thorough Testing.*—The standard naval battery of four pencil-and-paper group tests was given:

Test 1. Modified Shipley Abstraction.
2. Modified Bennett Mechanical Comprehension.
3a. Arithmetic. 3b. Mathematics.
4. Squares Test of Spatial Judgment.

The sum of scores on these four (with Test 1 doubled) gave the total score, known as T2, which was found to be the most useful index of all-round potentiality in the Navy. A test of dictation or spelling, and a mechanical + electrical information test, were generally applied, and when time allowed various tests were given experimental trials. No tests involving apparatus were feasible with such large numbers.

3. *Employment Interviewing.*—Though each recruit had received a tentative allocation at his recruiting centre, this could often be revised at the entry establishment in the light of the additional information collected, and of the Navy's current requirements. He therefore received a 10 to 30 minute interview with a P.S.O., leading up to a final employment recommendation. The interview was based on his original questionnaire, to which the results of the newly-taken tests had been added. Standardised oral tests of

fitting knowledge, machine tools, sheet metal work, petrol engines, diesel engines, and electrics, were applied in suitable cases, and other questions designed to elicit trade experience were often based on trade test pieces (cf. p. 149). Individual performance tests—Cube Construction or revised Kohs Blocks—were occasionally given. P.S.O.s possessed up-to-date information on the quotas for different branches and on job requirements. It should be noted that the procedure was as much one of guidance as of selection, and that the eventual choice, reached after integrating all the evidence, was as far as possible one which the recruit himself approved. Each P.S.O. was expected to conduct approximately 80 interviews a week, in addition to doing a share of the group testing and attending daily conferences on current work. But at rush periods P.S.O.s sometimes interviewed more than 30 men a day.

4. *Matching Supply and Demand.*—As indicated above, selection for a single branch is of comparatively little value when the total employable population is limited, since it almost inevitably leads to denuding other, possibly more important, branches of high-grade recruits. The psychologist must weigh up the demands of all the branches, and distribute men of the best quality fairly among them. Even those which require only a low average intellectual or educational level must receive a sufficient proportion of good material to yield potential N.C.O.s and instructors. Thus it was necessary to resist the temptation to allocate all the most able men as mechanics, writers, telegraphists, etc., leaving only the average and inferior men as seamen, since many seamen would later have to specialise in gunnery, torpedos, Radar or Asdic, and from their ranks the leading seamen, petty officers, and most R.N.V.R. executive officers would eventually be drawn.

Selection in the Forces was certainly more difficult than in industry or education owing to the constantly changing demands. Depending on operational needs, new categories would suddenly be started about which little was known, perhaps for security reasons. Other categories would close and large numbers of carefully chosen men would have to be transferred. Any stiffening in the training syllabuses for specialists might necessitate the raising of standards. Expansion of a category might involve lowering standards for a time in order to provide enough personnel.

5. *Re-allocation of Transfers.*—All men who became unfit for the branch to which they belonged, or who failed in their training

courses, were returned to H.M.S. *Royal Arthur* for re-allocation. In 1944 a scheme was instituted whereby the training schools had to supply details as to the reasons for failure. This was particularly useful since, when the failure was attributable to mistaken judgment on the part of a P.S.O., she could investigate the case herself and profit thereby. The morale of these failed men was often very low, hence the employment interview served the valuable function of re-establishing their self-respect, as well as discovering a more suitable job to which they could be transferred.

6. *Training and Research.*—Lastly, H.M.S. *Royal Arthur* served as a training ground for newly-promoted P.S.O.s, for neuropsychiatrists' assistants, and other selection personnel. Moreover, it provided ample opportunities for applying additional tests or for other investigations, whereas previously psychologists had scarcely been able to obtain more than one hour for testing.

Other Aspects of Naval Selection

Collaboration with Neuropsychiatrists.—Though there were many instances of fruitful collaboration between industrial psychologists and medical psychologists or neuropsychiatrists, the relationships of the two groups were less cordial than might have been anticipated. Some psychiatrists tended to be suspicious of the ability of psychologists to assess temperamental or personality qualities by interviews. Psychologists, on the other hand, were doubtful of psychiatrists' knowledge of jobs, and did not wish to become too closely associated, in the eyes of the Navy, with psychoanalysis and abnormal psychology.

Although it was proved, in joint investigations, that recruits picked by recruiting assistant Wrens as unstable were well worth referring to psychiatrists (cf. p. 157), various practical difficulties prevented such referrals from actually taking place. In H.M.S. *Royal Arthur*, however, some 7 per cent. of men suspected of neurotic tendencies by the P.S.O.s were seen by the resident psychiatrist, and 0·7 per cent. were recommended by him for discharge. He in his turn taught the P.S.O.s what to look for.

In their work at naval mental hospitals and clinics the neuropsychiatrists were provided with clinical assistants who were trained to apply portions of the Wechsler-Bellevue adult intelligence scale, and the Trist-Misselbrook revision of Kohs Blocks, to individual patients.

Work at Depots and Branches.—One or more P.S.O.s was attached to each of the larger naval holding and drafting centres, where they assisted in the re-allocation of recruits returned from sea. Many men who had entered the Navy before the S.P. department was established were thus tested and interviewed. The choice of seamen for training in gunnery, torpedo, Radar or Asdic was made at the depots, and, though first priority was always given to men who had been recommended at sea, the P.S.O.s helped considerably in raising and maintaining the quality of other trainees. In practice, almost the whole of the upper half of the naval population (i.e. men with good education and work record, not merely good intelligence) were recommended for some form of specialist training. The same P.S.O.s were largely responsible for collecting from specialist schools the results of training courses, upon which most of the follow-up investigations were based.

Officer Selection.—Experimental application of tests and psychological interviews in H.M.S. *King Alfred*—the training establishment for R.N.V.R. cadets—gave promising results in 1942. After study of the War Office Selection Board procedures (cf. Chapter IV), a modified scheme was introduced in the following year which spread out the selection process over a much longer period. Many possible candidates were ear-marked by P.S.O.s in entry establishments, and these were further interviewed by a psychologist and a R.N. executive officer. Those picked out underwent their first naval training in a new establishment, to which a psychologist and a testing officer familiar with army methods were attached. Some of the "leaderless group" and other army tests, adapted to naval conditions, were applied, but the Board's eventual decision as to whether the men should go forward as "C.W. candidates" was based mainly on observation of their behaviour during three months' strenuous training. Though there was no psychiatric interview, the psychologist studied each man intensively and applied certain projection and other personality tests.

"New-type" Boards were later introduced for candidates from the Fleet, and for "Special Entry" cadets. The candidates usually spent a week under observation, and were tested and interviewed. Thus the Admiralty Selection Board received a large amount of information on which to base its decisions. Work was also done in selection for engineering commissions, and with officers needed for fighter-direction. Some 10,000 candidates in all were covered during the war.

Although the standard tests, making up T2, differentiated well at the officer level, several other more difficult tests were tried out, the results with which are described in Chapter XIII.

Miscellaneous Fields of Selection.—Fleet Air Arm pilots and observers were selected and trained by the R.A.F. But telegraphist air-gunners mostly passed through H.M.S. *Royal Arthur*, and all the various types of air fitters and air mechanics (electrical, air-frames, engine and ordnance) were handled by S.P. department, constituting indeed one of its major—and most successfully accomplished—tasks (cf. p. 121). Selection and re-allocation were also applied on a large scale in the Royal Marines. The transference of large groups of surplus R.A.F. and army recruits into the Navy, and vice versa, raised serious problems of morale. Talks by psychologists and careful individual selection helped considerably to lessen the recruits' sense of frustration. Some help was given, mainly since the close of the war, in allocating Special Service recruits and Continuous Service boys. Among boys being trained as electrical, engine room, ordnance or air artificers (who had been selected by educational examinations) it was shown that a battery of mechanical and intelligence tests was of value both in the initial acceptance and in the allocation to the different trades.

Records and Headquarters' Activities.—During the war some 350,000 recruits passed through the selection procedure outlined above. All questionnaires were eventually housed in a central file, though a copy of the main items was first made and attached to each recruit's Service documents. Weekly returns of all test results and allocations were sent in. The filing system was so arranged that details on every recruit examined could be looked up at once, and it was usual for P.S.O.s in depots or other establishments to telephone the names and official numbers of individuals with whom they had to deal and to obtain the information forthwith. Two great advantages of this system were, first, that it was unnecessary to re-test recruits since first scores could always be traced, and, secondly, that it made possible extensive follow-up. The final course marks of, say, radio mechanics could readily be compared with the selection data collected one-and-a-half years earlier. Frequent returns from the specialist schools were also analysed at headquarters in order to provide a check on the quality of the material which was being sent to them, or on any changes in their requirements which might affect selection procedure.

The other main activities of headquarters staff included:

(i) The preparation and dispatch of questionnaire forms, test material, information bulletins, posters, etc.

(ii) Validation of selection procedures and other research.

(iii) Preparation of job analyses, often with the aid of specialist officers lent by the branches concerned. Information regarding training syllabuses and job descriptions naturally involved visits to training schools or to ships. But entry requirements, pay, promotion, etc., were better investigated at the Admiralty.

(iv) Liaison with other departments, such as the Directors of Naval Recruiting, Naval Training, Personal and Medical Services, and the Education Department.

Psychological Studies of Training

While selection was the Senior Psychologist's primary concern, his department was also authorised to undertake other psychological investigations. One of the major fields was training, and detailed enquiries were made into the selection and training of Asdic operators, Radar operators and radio mechanics, telegraphist air gunners, torpedo ratings and electrical mechanics and, in the engineering branch, of artificer apprentices, mechanicians and leading stokers. An elementary manual of psychological principles of instruction was published (Vernon, 1942), and a critical analysis of methods of learning to receive and send Morse code, based largely on recent American researches, was prepared. In these specialist courses the instructors and lecturers were carefully chosen and—at least towards the end of the war—many of them had been trained in instructional technique. Thus the psychologist's reports were less concerned with their competence than with:

(1) Success of trainees obtained from various sources, with varying previous experience and background.

(2) Propaganda for attracting suitable candidates.

(3) Organisation of the syllabus: working hours, distribution of periods, size of classes, etc.

(4) Effects of conditions of work and training on morale.

(5) Functions of theoretical instruction and its integration with practical training.

(6) Use of visual aids including diagrams, charts, films and demonstration models.

(7) Methods of examining and marking.

Attempts were made also to consider how far the "training as given," and examinations at the end of training, correlated with the "job as done," since it appeared in some branches that undue stress was laid on knowledge or skills which would seldom be required at sea (cf. p. 108). Probably most of the psychologist's recommendations did not carry much weight at the time. Nevertheless a large measure of goodwill was built up, and by the end of the war there was a readiness to consult psychologists in the early stages of planning instruction or of producing training devices, which would have been unthinkable five years earlier. Though psychologists were members of Admiralty committees on training apparatus and training aids, there was little opportunity for controlled experimentation. One large-scale investigation, however, demonstrated the value of film and filmstrip in the elementary instruction of seamen (Vernon, 1946a). Little use was made by naval schools of objective examinations until, towards the end of the war, American practice and the advice of British psychologists began to have some influence. A simplified manual of "new-type" examining was drawn up recently.

Another field where there was scope for psychological advice was documentation. Even before the war the N.I.I.P. had been consulted when the form for reporting on the efficiency of naval officers was undergoing revision. Recently a survey on behalf of the Director of Naval Recruiting showed how the paper work involved in the entry of recruits to the Navy (forms, records and returns) might be approximately halved.

During job analysis and training investigations many instances of bad layout or design of equipment were noted. Although such matters have largely concerned psychologists in industry, there was no ready means of bringing the psychological viewpoint to the attention of the naval scientists and engineers responsible for the designs. However, the Royal Naval Personnel Research Committee of the Medical Research Council did useful work in bringing together medical specialists and representatives of Admiralty departments dealing with training, clothing, gunnery, scientific research, naval construction, and others. Under its auspices investigations were made not only into equipment, but also into working and living conditions in submarines and the small craft of coastal forces, in the Arctic and in the Tropics (Critchley, 1947); into problems of night vision, colour blindness, seasickness, and other physiological and psychological matters (*Brit. Med. Bull.*, 1947). Further, with the co-operation of the War Office time and motion study psychologists, a beginning was made into devising simplified gun drills, and the Admiralty now has a motion study unit of its own engaged on a wide variety of work-efficiency problems.

Finally, a few investigations were carried out with opinion surveys for gauging the attitudes of recruits towards various aspects of their living and working conditions. One study dealt with opinions regarding new schemes of interior decoration and lighting of ships, others with views on demobilisation problems in which it was found, incidentally, that the work of vocational psychologists was sufficiently appreciated for a large majority to favour the giving of psychological advice on post-war careers.

Conclusion

While the ratings and officer candidates who underwent selection procedure were, in general, strongly in favour of its impartiality and its careful consideration of the capacities and interests of each individual, senior officers sometimes tended to be more cautious. The layman is often frightened of people who are supposedly expert in reading character or in explaining human behaviour. He dislikes their inquisitiveness and their criticisms of old-established ways of doing things, also their frequent resort to statistical evidence. It was natural also that the use of women to report on suitability for male occupations, and even to recommend possible officer candidates, should come under fire. Gradually, however, as the war progressed, and the value of the psychologists', P.S.O.s' and recruiting assistants' work became manifest, these objections died down and many critics were converted. In 1946–7 far more requests for assistance both in familiar and in fresh fields were received than the depleted staff could possibly cope with.

CHAPTER III

OTHER RANK SELECTION IN THE ARMY AND A.T.S.

Abstract.—Early experiments in the testing of Army and A.T.S. recruits achieved only a limited success. In 1941 the Directorate for Selection of Personnel (D.S.P.) was established, with powers to test, interview and make employment recommendations before recruits were assigned to their Arms of the Service. The staff of the directorate was almost wholly military, and certain difficulties arising from the restriction of the functions of qualified psychologists are pointed out. The work of selection was carried out by personnel selection officers and sergeant testers during the recruits' primary training. This consideration of each individual's abilities and interests at an early stage in his army career had excellent effects upon morale. The testing, interviewing and other procedures, and the ear-marking of potential officers and tradesmen, are described, together with the corresponding methods in the A.T.S. Much work was done in the re-allocation of converted units, of misfits, and other groups at home and overseas. A detailed analysis of Army and A.T.S. employments was made in 1941–2 and kept up to date. The directorate was not concerned with applications of psychology outside selection, but training, motion study and other investigations conducted independently are briefly mentioned.

The Army had even greater difficulties than the Royal Navy or R.A.F. in meeting its skilled manpower requirements, since it attracted a smaller proportion of men of high intelligence and education, and, being unable to reject any except the medically unfit at recruiting centres, it had to take in a considerable proportion of very dull men. Recruits were posted on call-up by the Ministry of Labour straight to some Corps or Arm, and there was no ready means for subsequent re-allocation of the unsuitables. The complexity of modern war has greatly increased the need for specialists, and the more technical Arms in particular failed to get enough tradesmen or trainable men. Moreover, as the Beveridge enquiry showed in 1941, many highly skilled men were engaged on semi-skilled or unskilled jobs. The effects on instruction, and on the morale of misplaced soldiers, were extremely serious. It was noted, for example, that more neurotic breakdown occurred among recruits under training than in battle, and unsuitability of employment appeared to be one of the factors responsible (cf. Sutherland and Fitzpatrick, 1945).

In 1939 the Army Council agreed to have an intelligence test given experimentally in military training units, under the technical direction of a small group of civilian psychologists, headed by E. Farmer. They devised a 20-minute omnibus verbal and mechanical comprehension test known as F.H.3 (later F.H.R.), which for the most part had to be applied by regimental officers who were unskilled in testing. Nominal rolls of low scorers were drawn up to which the Army psychiatrists occasionally had access. But as there was still no effective provision for transfer of men to other Arms, or for the discharge of mentally defective or unstable recruits, very little was accomplished.

An alternative scheme, put forward by A. Rodger early in 1941, with the help of G. R. Hargreaves, was the foundation of more fruitful developments. Hargreaves, an Army psychiatrist, had already obtained promising results with the Progressive Matrices test among R.A.M.C. personnel, and he influenced General Sir Ronald Adam, whose personal interest and energy, when he became Adjutant-General, played a major part in what followed. Eventually an advisory committee of psychologists was appointed, consisting of Myers, Drever, Burt and Philpott. Their main recommendations were that a new directorate for selection of personnel be set up, that an intelligence test be given at recruiting centres on medical examination, and that more detailed psychological and psychiatric examination be carried out at mobilisation depots, *before* recruits were allocated to any Arm (cf. Myers, 1942–3). These proposals were implemented in July, 1941, though it was not until a year later that the third one—namely, the General Service scheme—could be put into effect. Psychological staff had to be collected, personnel to carry out the programme had to be trained, a job analysis made of the main Army and A.T.S. employments, and suitable tests and procedures devised, before effective selection could be instituted (cf. Tuck, 1946; Directorate for Selection of Personnel, 1947).

The Staff of the Directorate

All executive functions were vested in the hands of Army officers, partly because of the small numbers of trained psychologists available, but mainly because it was thought that only a scheme run by soldiers would be acceptable to soldiers. The director, a brigadier, had to keep in intimate touch with the constantly changing requirements of the Army. Under him was a large staff of non-technical officers responsible for organisation, administration, supplies, etc., and all the day-to-day application of selection was carried out by personnel selection officers (mostly regimental majors or captains), presidents (colonels) and military testing officers at selection boards, and testing assistants (sergeants) who were given a small amount of psychological training. Psychologists, all but two of whom were put into uniform, were confined to planning and inspection of selection procedure, training the non-technical staff and carrying out research and development investigations. At the peak period there were 19 psychologists (including 5 women) and 31 officers or promoted sergeants with some training (5 women), together with nearly 600 non-technical officers and 700 N.C.O.s (of whom about 50 and 200 respectively were women). In addition a few qualified personnel were attached to the Directorate of Biological Research, the Department of the Scientific Adviser to the Army Council, and the Director-General of Army Medical Services (cf. Expert Committee, 1947).

Most of the trained psychologists were brought in from the universities; but the chief psychologists to the Army and A.T.S. and four others were past or present members of the N.I.I.P. staff. P.S.O.s, M.T.O.s and sergeant testers were volunteers from the Army, chosen in the first place for good intelligence scores, relevant past experience, and apparently suitable personalities. Sergeants had two or three weeks' training which dealt with the nature of the tests, their application and scoring. P.S.O.s received a month's training in principles of vocational selection, job analysis, assessment of educational level and occupational experience, Army requirements and documentation, the nature and use of tests and interviewing technique. A stiff written and practical examination was set before they were finally accepted, the former covering knowledge of procedures. The latter consisted of interviewing "stooges," obtaining information from them and making employment recommendations. Little or no instruction was given in general psychology as such, though reading was encouraged. Headquarters administrative staff and board presidents had no systematic training (the higher the rank and the greater the responsibility, the more difficult it became to demand technical qualifications). Nevertheless, many of them voluntarily undertook considerable study of the psychological background of their work and of the techniques they had to apply.

While the huge size of the task to be done in the Army necessitated a large non-technical staff, it seems doubtful, judging by the experience of the Navy, whether quite so much timidity about the reactions of the Army to psychology was justified. Without belittling in any way the very fine work done by all types and grades of staff, it must be admitted that the organisation did not always function smoothly. Psychologists suffered considerable frustration. They were commonly of lower rank than their non-technical colleagues. Policies which they advocated as scientifically sound were often rejected, and the methods they devised were often misapplied and misinterpreted by insufficiently trained personnel. Their training had perhaps predisposed them to seek what was best for the interests of the individual soldiers with whom they had to deal, and they were less experienced in envisaging the broader needs of the Army as a whole. Again, the fact that they were immured in headquarters (apart from occasional visits for inspection or for carrying out experimental investigations) tended to widen the gulf between the technical and the practical aspects of selection. The lot of the sergeant testers was particularly deplorable. Many were highly intelligent teachers and university graduates, but, except for a few brought to headquarters as research assistants and statisticians, they were restricted to routine application and scoring of tests under the command of P.S.O.s whose educational and psychological qualifications were sometimes inferior to their own, and were liable to be put on to cutting the grass or other duties at the whim of any C.O. A quarter to a third of them eventually achieved commissions by the same route as other Army recruits, but not on grounds of technical competence at their work. Presumably the lesson to be drawn is that psychologists cannot expect a complex institution like the Army to accept novel procedures merely on scientific grounds, that gradual education and infiltration rather than the imposition of technically valid methods are needed.

The General Service Scheme

The introduction of the Matrices test, given by sergeants at recruiting centres, allowed a better distribution of the available good-quality material among the Arms that needed it most, although no recruits were rejected by the test. When the General Service scheme came into operation, recruiting-centre testing was abandoned (except for regular volunteers). During the winter and spring of 1941–2 there was intensive work on an Army job analysis, the preparation of new tests and experimental trials of the full-scale procedure. Under the new scheme all recruits were enlisted into the General Service Corps and sent to one of the numerous primary training centres (P.T.C.s) up and down the country for six weeks, during which selection took place and a common syllabus of initial training was given. Some 160 P.S.O.s and 500 sergeants dealt with the fortnightly intakes which normally included some 12,000 men, but occasionally rose to double this number. The following tests were given:

Progressive Matrices.
Test 2. Bennett Mechanical Comprehension.
Test 3A. Army Arithmetic and Mathematics.
Test 4. Squares test of Spatial Judgment (dropped in 1943).
Test 25. Verbal test (initially Test 17. Messages).
Test 16. Physical Agility.

Additional confirmatory tests were applied to those considered suitable for certain types of employment, namely:

Test 12 or 21. Clerical or Instructions (later introduced into the general battery in place of Squares).
Test 8. Assembly test of Mechanical Ability.
Tests 10 and 19. Morse Aptitude and Auditory Acuity.

Scores on each test were expressed as Selection Grades or S.G.s (cf. p. 177). After follow-up investigations had revealed the value of the Matrices, Bennett, Arithmetic, Verbal and Instructions tests in most Army jobs, the S.G.s on these tests were summed to yield the measure of general intelligence known as Summed S.G.

A qualification form (cf. p. 132), similar to the naval biographical questionnaire, was filled in under supervision of the tester. This was used as the basis for the interview of each man by the P.S.O., in which the details were clarified, the recruit's interests consulted, and employment recommendations reached. Most of these recommendations consisted only of broad categories, both because sorting had to be carried out centrally, and because supply could not otherwise be matched with demand for all the tremendous variety of Army jobs. The following list of T.R.s (Training Recommendations) was used, after some revision.

0. Potential tradesmen.
1. Drivers.
2. Mechanical maintenance (later omitted).
3. Signallers.
4. Constructional, building, and other heavy duties (originally layers and fine operators).

5a. Clerical and 5b. Storekeeping.

6. Mobile and combatant duties, including riflemen and gun numbers.
7. Orderly duties.

8. and 9. Pioneer Corps or discharge.

Three recommendations were given in order of suitability, though often a recruit received two identical T.R.s, e.g. 667, meaning that both his first and second choices were for straightforward combatant jobs. The T.R.s and other information such as age, medical category and Matrices (later Summed S.G.) results were forwarded to the War Office and entered on punched cards for sorting and posting. If the requisite number of, say, drivers for R.A.C., infantry and other Arms could not be obtained from men with T.R.1 first choices, second choices were drawn upon. The postings were then sent back to the P.T.C. within three weeks, and the P.S.O.s entered the final T.R., this time with a suggestion as to a specific job within the T.R. (e.g. dispatch rider, tank driver, etc.).

In making their recommendations P.S.O.s had, in addition to Army job analyses, lists of minimum standards on the various selection tests, medical categories and other qualifications regarded as essential for each T.R. Some perhaps applied this scheme too mechanically, while others were too apt to ignore it and trust their own subjective judgments. But the majority certainly attempted to integrate all the relevant information about each man, including his test results as one important but not overwhelming factor. The P.S.O.s' work and the sergeants' testing and scoring were supervised by Command P.S.O.s who were psychologically qualified officers. Moreover, records of the allocations were forwarded to headquarters where the more glaring misplacements or disregard of instructions were checked. The occasional P.S.O.s who were clearly incompetent could thus be eliminated from the directorate.

Some 14 per cent. of recruits were referred to an Army psychiatrist working at the P.T.C., including all those of very low intelligence, men considered to be poor in "combatant temperament," and possible neurotics or psychopaths. Many of these were passed as normal, but those with psychiatric disabilities were either allocated to appropriate non-combatant employments, or to the Pioneer Corps (armed or unarmed), or sent to mental hospitals or discharged. By wise placement many emotionally unstable men were helped to better adjustment than they showed in civil life, and the Army's use of mentally defective and very dull men in the Pioneer Corps was an outstanding success. Recruits who as civilians had been drifters or unemployables were found to give excellent service at simple labouring tasks, when working and living with others of a similar level of ability under specially chosen officers and N.C.O.s. Their morale was high, and their sickness and delinquency rates low. Moreover, the dull, hard jobs of the Army were probably better done than they would have been by more intelligent recruits (cf. Rees, 1946).

Potential tradesmen were also identified individually. P.S.O.s were provided with a *Guide to Civilian Occupations* which helped them to assess both the amount of trade experience claimed by recruits and its relevance to Army trades. Later, a series of written tests of trade knowledge was prepared, each covering a particular Army trade. Often the number of skilled, or semi-qualified men was insufficient; and the balance was made up from inexperienced recruits of high intelligence who were considered to be rapidly trainable. A "Short List" was issued each fortnight of the trades for which candidates were specially needed. Actually, as shown later (p. 120), these men put up by P.S.O.s did distinctly better at trade courses than the semi-qualified tradesmen submitted by the Ministry of Labour, or others put up at their own request, or by their C.O.s.

P.S.O.s also had the task of assessing good or poor employment record, very strong or weak combatancy or aggressiveness (some 5 per cent. of men were usually placed in each extreme category), and suitability as a potential officer. Those in the latter group were automatically sent to War Office selection boards after six months' training, along with other candidates recommended from their units, and, in fact, provided the major source of officer material in the second half of the war. When strong experimental evidence had accrued as to the value of certain of the standard tests in selecting successful candidates, P.S.O.s were instructed to take particular account of the combined score on these tests*. Many however, preferred to rely on their own hunches regarding the officer-like qualities of their interviewees.

Between 1942–5 some 700,000 recruits passed through the G.S. Corps of whom 9 per cent. were chosen as potential tradesmen and 6 per cent. marked as potential officers. Each man's qualification form, with test results and T.R.s entered, was attached to his Service documents, but a copy of the main items was entered on a card and stored in headquarters for research purposes. Unlike the naval system, there was no easy means of disinterring the selection information about any given individual, and when recruits came up for later re-allocation, their original scores were seldom available. Many men went through the same battery two or more times —a proceeding whose bad effects were only partially overcome by the provision of retest norms. Not until the end of the war was it possible to prepare parallel versions of the chief tests.

Although less stress was laid in the Army than in the Navy on self-guidance, P.S.O.s gave an introductory talk to each batch of recruits and spent part of their interviews providing information on the types of employment likely to suit them.

In conclusion, selection procedure on so large a scale had its inevitable weaknesses, but there is no doubt that the fair and thorough consideration given to each individual had a most favourable effect on morale. Even apart from the saving in misplaced manpower, of which evidence is given below, this feature provided justification for its introduction into the Army.

Initial Selection in the A.T.S.

Selection procedure in the A.T.S. which is fully described in an article by Mercer (1945), differed little from that in the Army and was run under the same directorate. By 1942, with the introduction of compulsory national service for women, the numbers in the Service reached over 200,000, and there were more than a hundred different employments. Some of these could be filled directly by recruits engaged in similar civilian jobs, but there were many with no civilian counterpart, such as instrument numbers (anti-aircraft personnel), wireless operators, tinsmiths and fitters. Again many different types of work might be involved in jobs called by the same name; for example, there were eighteen kinds of clerk. One of the first steps, therefore, as in the Army, was a job analysis of the main varieties of employment, special attention being paid to the conditions under which the work would be done (sedentary, outdoor, dirty, long hours, etc.), and the degree of individual responsibility involved, as well as to the capacities and training needed.

---

* The combination was score on Instructions + Bennett, minus the score on Arithmetic + Matrices. This closely duplicated the weightings obtained from the multiple regression equation (cf. p. 104) between the G.S. test battery and W.O.S.B. Pass or Fail.

At the beginning of the war most of the allocation was done by the commandants of the basic training centres who often had to deal as best they could with, perhaps, 250 recruits in an afternoon. Early in 1941 an experimental battery of sensory-motor and perceptual tests, devised by the Cambridge Psychological Laboratory, was instituted for the selection of A.A. instrument numbers, This was taken over by the directorate, but later found to be less reliable and valid than ordinary paper-and-pencil tests (cf. p. 248). The Matrices test was given at recruiting centres by A.T.S. sergeant testers from 1942 onwards, and between 5 per cent. and 25 per cent. of the lowest scorers were rejected, depending on the number and level of recruits required. It should be noted that there was no temptation to fail this test in order to escape conscription. Most recruits were volunteers, and all the others had chosen the A.T.S. in preference to some alternative form of service. There was no interviewing or allocation at recruiting centres.

At the A.T.S. basic training centres, the procedure was very similar to that in Army P.T.C.s, except that, with the smaller numbers, recommendations for particular jobs could be made instead of for general types of duty. Usually two or three suggestions were listed, which enabled supply to be matched with demand. P.S.O.s in the A.T.S. were chosen to be as tactful and sympathetic in manner as possible, since many personal problems were apt to be raised by recruits. Referrals to (women) psychiatrists were of value, but were less frequent than in the Army, both because of the exclusion of low Matrices scorers and because undue lack of combatancy did not need to be catered for. A large amount of information on jobs was made available to recruits by means of photographs, films and talks. With the exception of the Agility test, almost the same battery of tests was used, but a simple form of computation-checking test was substituted for the Arithmetic 3A test (cf. p. 229), a spelling test was added, and standardised stenography and typewriting tests were available for those claiming clerical experience. Nearly 40,000 recruits passed through this procedure of whom, it was subsequently found, 94 per cent. were successful in the training to which they had been assigned.

Re-allocation and Conversions

Probably as much work was done on re-allocation or transference of men and women to new jobs as on initial selection. During 1942–3 some hundred battalions were converted from Infantry to other Arms. Later, when the danger of air-raids decreased, most of A.T.S. anti-aircraft employment was closed down. Casualties in the armies overseas, or medically down-graded men might need new employment. Many units formed before the institution of selection procedure contained numbers of misfits who, having failed at the first job to which they had been posted, were put on to orderly duties or had nothing to do. Recruits of long service in the A.T.S. (enlisted before 1942) could be recommended by their C.O.s for up-grading to a more skilled and better paid job. Finally, some new need might arise, particularly in overseas units for which suitable men had to be rapidly collected from other jobs, as when the Army of the Rhine urgently demanded deep-sea divers for work at Amsterdam Harbour, and the demand—which would normally have taken weeks of correspondence with the units—was instantly met by the D.S.P. documentation scheme.

Some of the more straightforward re-sorting could be done by Command P.S.O.s or headquarters staff merely through scrutinising the qualification forms of recruits who had been through General Service. For large-scale schemes, personnel selection teams visited the units and applied the usual testing and interviewing procedure. Such teams were also set up in Egypt, North Africa, Italy, etc., to deal with local problems. In 21st Army Group records were collected before D-Day on Findex cards, to facilitate selection for special duties and reorganisation or regrouping after battle. Every N.C.O. and other rank, in addition to being tested, was rated on a number of qualities pertaining to efficiency at a conference held between a P.S.O. and officers or N.C.O.s who knew the men best. The same technique was of value in disposing of the 35,000 A.T.S. in A.A. units.

Perhaps the most noteworthy re-allocation scheme, from the psychological angle, was that of the Army and A.T.S. selection centres, where misfits, down-graded recruits and the like were sent. All were re-tested, medically examined, and very thoroughly interviewed by P.S.O.s, and a third to a half were seen by psychiatrists. A conference was held between the C.O., the medical officer, psychiatrist and P.S.O. at which final decisions were taken, about 30 recruits being covered in 2–2½ hours. Those for whom no appropriate employment was available could be, and were, discharged. This system was particularly valuable in raising morale, and substantially similar methods were used for dealing with escaped and repatriated prisoners of war. In a single year (1943–4), 30,000 to 40,000 men passed through A.S.C.s, and a similar number were re-allocated overseas; 75,000 returned prisoners were dealt with during 1945.

Miscellaneous Psychological Activities

The technical staff at headquarters engaged on routine statistics and records, development and research, was much larger than in the Royal Navy. Some 20 technical officers, P.S.O.s and psychologists and 50 sergeant assistants were often involved. Test construction and follow-up are described in later chapters. Job analysis was continually being extended to cover the smaller categories omitted in the first survey, and to allow for any alterations in the nature of employments. The W.O.S.B.s and A.T.S. had their own job analysis teams. The usual method was for an officer specialising in an employment to collaborate with the technical officer. The two visited units to study the recruits at their training or on the job. Towards the end of the war, simplified editions of analyses of all the jobs in certain Arms were issued which aimed to convey to P.S.O.s, by means of diagrams, drawings and photographs, as vivid a picture as possible of the work and the conditions under which it was done. The minimum standards or other qualifications issued for each employment also underwent continuous revision. An additional P.S.O. was attached to each Arm in order to keep D.S.P. in constant touch with its selection problems, and to advise on any changes.

One branch of D.S.P. was concerned solely with exceptional recruits who required special employment recommendations. Some of these were psychoneurotics; others were men of very high intelligence, but with no signs of officer quality or of mechanical aptitude—for example, university graduates in languages or classics, for whom posts as interpreters, script writers, radio announcers, etc., could often be found. Another activity was dealing with "atrocity stories," i.e., complaints by senior officers or politicians of gross misplacements, or rejections of particular recruits.

Much assistance was given to psychologists in the Allied Forces. Thus, the Dutch, Belgian and French Armies set up similar selection organisations and translated several of the S.P. tests. Several D.S.P. teams were sent abroad to study and carry out selection of non-English speaking recruits—Indians, Palestinians, East and West Africans, etc.

Boy tradesmen are recruited for the regular Army, much as for the Navy, by academic examinations at about the age of 14, and receive some four years' apprenticeship training. Here, too, the value of mechanical tests and of an interview by an industrial psychologist were demonstrated (cf. p. 245).

The directorate was specifically established to deal with selection. This had the advantage that it could prohibit units from applying home-made tests. For example, one mechanic's branch was in the habit of throwing out numerous promising mechanics on the basis of an unstandardised general knowledge examination. But it also meant that there was no brief for carrying out other psychological work. Several researches were, in fact, undertaken as being relevant to selection; for example, the effects of menstruation on A.T.S. test performances, the effects of attendance at physical development centres or at basic education courses, or of membership of pre-Service organisations (cf. Chapter XI). But training investigations in particular fell outside the directorate's scope.

Two memoranda on methods of Army instruction were drawn up by Burt and Valentine (1942, 1943) on the basis of replies to questionnaires sent to past students who were serving. The main defects mentioned were:

(1) Mechanical, parrot-like teaching.
(2) Unnecessary enumeration of parts of weapons at early stages
(3) Use of technical vocabulary and unfamiliar words.

(4) Too much crowding of instruction; bad organisation of courses.
(5) Lack of "learning by doing"; lack of appeal to interest.
(6) Lack of visual aids, and of elementary textbooks.
(7) Use of sarcasm or abuse by instructors; unsympathetic handling.
(8) Inadequate selection of N.C.O.s suited to instruction; lack of training in instructional technique.
(9) No homogeneous grouping of classes in ability.
(10) Haphazard and subjective examination techniques.

Later, W. Stephenson, who was appointed Consultant Psychologist to the Army Medical Services, carried out extensive time studies and other investigations of instruction, recording for example, the relative times devoted in specimen periods to demonstration, to practice and to oral instruction, and the use of visual aids. Several reforms arose out of such enquiries, and "methods of instruction teams" were trained which toured the units and Army schools.

A time and motion study group, headed by Ungerson (1945), was established under the Scientific Adviser to the Army Council, which recorded existing drills for gun teams and drew up revised and more efficient methods. For example, the rate of fire of a coastal artillery gun was doubled, the time for unloading a jeep and trailer from a glider was reduced from ten to two minutes, and an improved scheme of handling ammunition in a tank was worked out. Other problems with psychological bearings which were tackled by the Medical Research Council, the Directorate of Biological Research, or the Directorate of Army Psychiatry are mentioned in the Expert Committee's Report (1947). These included design and layout of equipment and of operations-rooms controls, effects of poor ventilation on motor co-ordination, effects of certain drugs, and morale and disciplinary studies.

CHAPTER IV

ARMY OFFICER SELECTION

Abstract.—The old-type interview boards for the selection of officers showed considerable shortcomings by 1941, and new War Office Selection Boards were set up, each staffed by a president, a psychiatrist, military testing officers and a psychologist or sergeant testers. These conducted much more thorough two- or three-day examinations of candidates and—whether their methods were entirely technically sound or not—they won the confidence of the Army and stimulated a continuous flow of good material. The functions of the various board members are discussed in some detail, also certain conflicts which arose between the military and psychological viewpoints. Psychologists were chiefly confined to training, research, and the application of intelligence and projection tests to candidates. The diagnosis of the character qualities which are so important to an officer was mainly done by psychiatric interview, and by the "leaderless group" and other practical exercises. These were designed not so much to reveal officer traits (e.g. leadership) or abilities, as to bring out the candidates' social reactions under conditions of strain. Some criticism is made of the scientific value of these diagnostic methods, but they nevertheless have important bearings on the selection of executives, civil servants, and other high-grade personnel in peace-time.

In the first two years of the war, the selection of temporary officers was based on recommendations from C.O.s and quarter- to half-hour interviews of the candidates by boards of senior officers. The system worked fairly effectively so long as there was a large supply of good material, e.g. from the public schools. But when this source began to dry up, the boards, being faced with recruits whose social and educational backgrounds were entirely unfamiliar, were unable to discriminate effectively. Unsuitable candidates were often passed and sent to O.C.T.U. (Officer Cadet Training Units), where large proportions failed, with unfortunate effects on the morale of the remainder. Moreover, so many candidates who might have succeeded were rejected by the boards, often—according to their own accounts—on flimsy grounds such as Grammar School education or socialist opinions, that recruits lost confidence in the system, and there was a real danger of insufficient officers being forthcoming. Again, there was less opportunity than in 1914–18 for selection on the basis of performance in battle.

Early in 1941 experiments were carried out by two psychiatrists attached to Scottish Command, encouraged by the G.O.C., Sir Andrew Thorne, who had previously been military attaché at Berlin and had observed some of the elaborate selection techniques developed by German military psychologists (cf. Chapter I). Officers attending courses at the Edinburgh Company Commanders School were given psychiatric interviews and intelligence tests, together with other tests on the German model. While the latter gave unpromising results, the correspondence between the psychiatric diagnoses and the school's estimates of the officers' worth was very striking (cf. p. 160). By 1942 the first experimental "new-type" War Office Selection Board has been set up. The methods worked out here were adopted by other new boards, a dozen of which were started in various parts of the country by October, 1942. A.T.S. Officer Selection Boards followed in 1943. Later boards, on the same lines, were attached to Armies in the Middle East, India, Italy and Western Europe. By the end of the war some 140,000 candidates had been through the new procedure, of whom about 60,000 passed.

So many popular descriptions have been published that a brief outline of the procedure will suffice (cf. Garforth, 1945; Directorate for Selection of Personnel, 1947). Most boards dealt with 120 or with 64 candidates a week. Though there were many variations between boards it was usual for candidates to visit for three days. During this time they filled in a biographical and a medical questionnaire, and were given certain intelligence and personality tests, the latter being of the projection type. A proportion were interviewed by a psychiatrist, and all were interviewed by the president or deputy president—senior Army officers. Each batch of eight to ten was under the charge of a military testing officer (M.T.O.), who messed with them, and did his best to put them at ease and to observe their natural social behaviour. He too applied the series of practical tests such as group discussions and lecturettes, command situations, obstacle courses and "leaderless group" tests designed to bring out the candidates' initiative, co-operativeness, leadership and other social qualities. Often a visiting officer, with the president and other members of the board, observed some of these tests. To candidates for technical Arms such as R.E. and R.E.M.E., a mathematics paper might be given and a further interview with a specialist officer to enquire into their engineering or other experience. At the end of the period a conference was held between all members of the board where the judgments of M.T.O., psychiatrist, technical officer if any, and president were compared and discussed, and doubtful cases were considered in detail before the president gave his final decision.

We can best study the method more fully under the headings—the functions of the president, psychiatrist, psychologist and M.T.O.*.

The President

At W.O.S.B.s the problem of the Army versus the technician was intensified. The new methods raised far more controversy and opposition than did any other aspect of personnel selection. As in the Navy, criticism came from regular officers rather than from candidates. It was clear that the latter considered the new boards a vast improvement on the old, for applications for commissions greatly increased in number. Time and again candidates—even failed ones—commented on the fairness, friendliness and thoroughness of the scheme. They even resented having to forgo the psychiatric interview, though this was the feature that aroused most suspicion. Any candidate who did think himself unfairly treated was at liberty to apply again and to come before the same, or another, board, which took no account of his previous failure. We do not wish to imply that Army officers were biased. Their opposition was very natural, since it was they, as commanding officers, who in battle had to entrust the lives of their men to the junior officers accepted by the boards. Once they had seen the scheme working, the great majority were converted in its favour.

For these reasons, it was thought that no scheme entirely run by technicians could prove successful. As Sutherland and Fitzpatrick (1945) put it, this would have meant introducing "a foreign body into the tissues of the Army." Though the aim was to educate the Army gradually into accepting scientific methods, the compromise eventually achieved showed considerable technical defects. Psychiatrists, psychologists and M.T.O.s were technical advisers to the president; and each president could run his board as he wished, with as much or as little reference to the technicians as he wished, subject only to the controlling authority of the Director for Selection of Personnel, himself a professional soldier. Hence, the president, representing the Army, was responsible for the final decisions; hence also a major part was played by M.T.O.s who were regimental officers. This meant considerable dependence on the subjective judgment of a single man, and considerable divergence between different boards. The president (or his deputy) did have a detailed questionnaire on educational, employment and Army history, and the C.O.'s opinion, as a basis for his half-hour interview, but no standard scheme of interviewing was followed. Chiefly he attempted to gauge "officer quality" and "leadership" by enquiring into interests and past achievements, and the candidate's attitudes towards an officer's rôles and responsibilities. In addition, he tried to assess the relevance of the candidate's qualifications to his chosen Arm and could, if he thought fit, recommend his acceptance for some different Arm. Differences between boards greatly hampered all subsequent attempts to validate the procedures (cf. pp. 122–127), and led to significant discrepancies in pass rates and distributions of grades, even though these were considerably smaller than those found among old-type boards. Thus, in one experiment where strictly comparable groups of candidates were sent to two boards, the respective pass rates were 23 per cent. and 48 per cent. This, however, was probably an extreme case.

---

* The present writer is indebted to members of the staff of the W.O.S.B. Research and Training Centre for information, and to a paper on "The Growth and Development of War Office Selection Boards," read to the British Psychological Society, Scottish Branch, Oct. 1946, by A. O. Heard. The views expressed here are, however, his own.

The Psychologist

The chief technical officers consisted of five or six fully qualified psychiatrists and psychologists, some senior M.T.O.s, and several trained assistants. They were responsible for developing and improving the methods used by all the boards, but as their functions were advisory only, they had very little control. For the most part they were isolated in the research and training centre (R.T.C.), where they were not given any opportunity, until the end of the war, to carry out day-to-day selection themselves. They had little contact, also, with the technical staff of D.S.P. working at the War Office, since their approach—based mainly on psychopathology and on psychological "field theory" (cf. p. 61)—was so very different from that of headquarters psychologists—based more on industrial psychology and psychometry.

When the boards started, there was no possibility of finding fifteen or more psychologists qualified to undertake detailed diagnostic testing and interviewing, though later many of the more promising sergeants were commissioned and given increased responsibilities. At first then, apart from the psychological staff of R.T.C., board psychologists were sergeants, three of whom acted as assistants to each psychiatrist; they did not attend the final conferences. Another reason for the minor rôle of psychology was that officer-suitability appeared to be chiefly a matter of character and personality, and in the absence of objective tests of the desired qualities, interview techniques which psychiatrists themselves had evolved successfully at Edinburgh constituted the most promising approach.

The sergeants' functions were not, however, purely psychometric. They applied and scored intelligence tests, but also gave the projection tests and to some extent interpreted the results of these, with a view both to separating off the candidates on whom the psychiatrist could most profitably concentrate from those whom he need not interview, and to feeding him with information or "pointers" about the former.

Three twenty-minute group tests were used, namely, the Army V.I.T. (Verbal Intelligence Test)—an omnibus test of conventional type, and new and harder versions of the Matrices and Shipley Abstraction tests. Raw scores on these were converted into equivalent scores, the sum of which was then converted into an officer-intelligence rating. This rating ranged from 10 to 0, with 3 corresponding approximately to the median for the general Army population, and 7 to the median for officers. In general, no candidates with O.I.R. below 3 were accepted. On the other hand very little stress was laid on a high rating; candidates scoring 10 were not looked on any more favourably than those scoring, say, 6. Several other tests were available for special purposes such as the Wechsler, Mill Hill and other vocabulary tests, and the Trist-Misselbrook Kohs Blocks. Experiments were carried out with Weigl Sorting, Carl Hollow Square and other diagnostic tests, particularly in the early days of the Edinburgh Board, though no definitive results were obtained. An assembly test for engineering cadets was devised to yield qualitative ratings of mechanical interest and comprehension, methods of work, etc., rather than quantitative scores. A detailed information test on workshop procedures was also prepared for R.E.M.E. cadets.

In 1944–6 standardised educational achievement tests were constructed for such groups as schoolboy candidates for university short courses, which were intended to supply potential officers for technical Arms, and for candidates for Army College courses. These included objective examinations in arithmetic, algebra, geometry, calculus, heat, light, electricity, mechanics, general science, current affairs in history and in geography. The tests were of school certificate to higher certificate standard.

The first questionnaire has been mentioned above. Its primary aim was to indicate the candidate's opportunities and what he had made of them. The second asked for various details of medical and family history, and was seen only by the psychologists and psychiatrists. To some extent this acted like the personality inventories in which neurotics tend to check large numbers of symptoms (cf. p. 257). Candidates were also required to write two brief descriptions of themselves as seen first by a good friend, secondly by a severe critic. Sergeants were trained to extract from these documents the points of major psychological interest. In the two group projection tests, direct self-reference was intentionally avoided. A fifty-word free association test was given with the title "Test of Quickness of Imagination." The stimulus words were presented visually for fifteen seconds each, candidates being instructed to write down the first idea that came to mind, either words or sentences. Some of the words were originally chosen to have ambiguous meanings, e.g. "lead," it being hoped that the more suitable candidates would react to their military meanings. But this feature possessed no diagnostic worth.

Lastly, six of Murray's Thematic Apperception pictures were shown by slides for four minutes each. Candidates were instructed to write short stories describing the social situations suggested. Sergeants received training, both at R.T.C. and from their own psychiatrist, in the interpretation of responses to these projection tests, neither of which was objectively scored. On an average, half-an-hour was available for reading the questionnaires and projection responses and integrating their deductions into a summary of the salient personality features, according to a more or less standard scheme. Note was made of the frequency of appearance of such attitudes as anti-social, disruptive or cohesive, also of undue morbidity, or of special types of symbolism. Degree of self-insight, breadth of outlook, and extent of identification with the Army's needs were assessed. Admittedly the technique was largely intuitive, and, although investigations into reliability and validity were planned, it was not possible to complete them. Nevertheless, a moderate degree of consistency was found between different sergeants judging the same material. Moreover, when a large number of psychiatrists and psychologists independently interpreted the pointer material of 38 candidates, the promising correlation (cf. p. 104) of 0.58 was obtained between their modal judgment and the final board grading.

With the provision of more commissioned psychologists towards the end of the war, it was arranged that candidates whose pointers were not followed up by psychiatrists, and whose personalities appeared less complex, should be interviewed by a psychologist, though his province was strictly limited to "surface" traits and sentiments. At the end of 1946 the shortage of staff became so acute that both psychiatrists and commissioned psychologists were withdrawn from the boards altogether.

The Psychiatrist

Since there was only one psychiatrist at most of the boards, and since his interviews lasted from 20 to 60 minutes, the proportion of candidates seen had to be limited. But an additional reason for this was the manifest desire of Army authorities to reduce his rôle to a minimum. While their suspicion was natural enough, it was certainly misguided. Only those who were ignorant of board procedures and of the nature of the discussions at the final conference could continue to believe that psychiatrists spend all their time smelling out sexual complexes. It is true that in occasional cases of superficially well-adjusted personality the psychiatrist might discover psychopathic or neurotic tendencies and recommend their rejection. But at least as often he drew attention to underlying potentialities and strengths in outwardly unimpressive or diffident personalities which other members of the conference had been inclined to reject. His services as a purely medical examiner were also sometimes needed. But the main value of the psychiatric approach, and the reason for its success in the early experiments may, perhaps, be expressed as follows.

The boards were concerned to select candidates who would:

(i) Pass their O.C.T.U. courses, and acquire the necessary technical proficiency.
(ii) Stand up well to the stresses of battle, showing resource, leadership, aggressiveness and caution when these qualities were needed.
(iii) Look after his men well and gain their confidence.
(iv) Co-operate effectively with other officers and make a good impression on his seniors.

Several other rôles might be mentioned, but these are among the chief ones. With most of the candidates little direct evidence was available, particularly regarding (ii) and (iii), thus the board had to predict future behaviour, often in young and immature personalities, from the indirect indications of past history. Now it is the medical psychologist—particularly the psychoanalyst—who has been chiefly responsible for introducing the viewpoint, generally accepted by contemporary psychologists, that our behaviour is never a random or chance reaction to the circumstances of the moment, but is always determined by or arises out of that organisation of innate and acquired tendencies which we call the total personality. Further, that we are not conscious of the true nature of many of the most important of these tendencies. Thus a more complete explanation of our present characteristics and a more accurate prediction of future trends is likely to be obtained by exploring the underlying mechanisms, and this the medical psychologist is particularly well fitted by his training to undertake, though others, such as the normal psychologist or the literary artist, are often capable of equally penetrating diagnoses. Having studied abnormal and disintegrated personalities the psychiatrist is quick to recognise the relatively well-balanced and integrated personality, and one of the principles upon which W.O.S.B. psychiatrists worked was that it is the harmoniously organised man who is most capable of showing aggressive or cautious, tactful, deferential, firm or sympathetic characteristics whenever these may be needed, and is least likely to give way under stress. Finally, it will be noted that most of the desirable officer qualities consist of appropriate reactions to people; there is little of the job which will not be carried out in a social context. Medical psychologists believe, rightly or wrongly, that adult social reactions are based largely on the attitudes to parents and other humans established in infancy and early childhood, hence the relevance of their enquiries into family history and childhood memories. Among A.T.S. officers the social side is if anything more important, since they are not usually required to be technical specialists and will seldom be subjected to operational stress. Thus the personality organisation and attitudes to people desirable in a woman officer (which may be very different from those in a man) was the subject of special study, largely by women psychiatrists.

The Military Testing Officer

The M.T.O. was originally introduced as a "cover plan"; he was to apply tests of a military nature which would impress the Army while the real job of selection was undertaken by the technicians. However, his functions were progressively clarified, and their value increased, through the investigations at the research and training centre. In the first experimental board he tried out some of the practical situation tests used in Germany, and improvised others. For example, he took his group of eight candidates to a stretch of country and there posed a series of tactical or other problems to candidates in turn. Each candidate again might be put in charge of the rest so that he could display his handling of them in military situations. Obviously, the success of candidates at such tests is likely to depend as much on their previous training and experience as on their personal resourcefulness or other qualities. Group athletic games and elaborate obstacle courses were also devised, with a view to bringing out co-operativeness, aggressiveness and "guts," quick judgment, endurance and the like; but physical status and bodily skills probably entered too largely. Lecturettes, given with or without preparation, and informal discussions were supposed to show fluency, confidence or social aplomb, together with acceptable or undesirable attitudes towards the topics discussed, e.g. discipline. M.T.O.s were usually men with battle experience, whose age was fairly close to that of the majority of candidates. Hence their personal experience of the work for which candidates were being selected was of value. But they had no systematic instruction in the assessment of personality or avoidance of halo. Although many of the tasks appealed to the candidates' sense of fitness and fairness—also to outside critics—they were quite unstandardised and unscoreable, and their diagnostic worth was extremely dubious.

In the view of the R.T.C. psychologists, the search for particular traits characteristic of officers, and for tests to measure these traits was likely to be a waste of time. Influenced by the field theory approach of Lewin, Moreno and other writers in America, they could not accept traits as predominantly constant qualities of an individual, existing independently of the context in which they are expressed. Successful officers do not all show the same traits; thus, it would seem to be the total configuration of traits in their personalities rather than the individual traits which makes for success. It follows that a candidate should not be thought of as possessing a certain amount of leadership which he can display both in test and in real life situations. His personality is an organised whole, a system of tensions or needs, which interacts dynamically with the varying demands of different situations. "Officer quality" should, therefore, be analysed in terms of the main rôles that future officers will be called on to play. By setting appropriate tasks a similar system of forces can be set up at the W.O.S.B., whose interplay can then be observed by the M.T.O. or other board member. The candidate's most important rôle will be that of leader of a small group, and he should be able to uphold his own position in such a group, to give the group direction, and at the same time maintain its cohesion or solidarity against internal or external disruptive forces.

It was from this theoretical background that the basic series of leaderless group tests was evolved in 1943 by Bion—tests which constituted the most original feature of W.O.S.B. procedure. In a leaderless group test, a group of, say, eight candidates as a whole is given a task, and left to work out its own solution. No leader is appointed, and the M.T.O. is an entirely neutral observer. Each candidate is out to distinguish himself, but the task is so designed that this is only possible through the success of the group. In other words he is forced to submerge himself and serve the group's interests. Each candidate resolves this conflict in his own individual manner. Some become mere "passengers" who give up their own desire to shine; others become "thrusters" who subordinate group interests to their own. More effective solutions, lying between these extremes, are shown by those who spontaneously attract the confidence of their fellows, and who come to the fore although working for the group's success. Thus the reactions of candidates to such situations are considered to be diagnostic of their behaviour towards their platoons under Active Service conditions. Bion (1946) writes: "The essence of the technique . . . was to provide a framework in which selecting officers, including a psychiatrist, could observe a man's capacity for maintaining personal relationships in a situation of strain that tempted him to disregard the interests of his fellows for the sake of his own."

The tasks were designed to be "real" ones which would bring out the candidates' natural behaviour, not artificial test situations. To the candidates or the uninitiated observer they appeared to pose a practical problem, but this, of course, concealed the underlying social problem in which the selection staff were interested. The basic series of tests, although run by the M.T.O., came to provide a common meeting place for all members of the board. They found that they could more effectively apply their particular techniques after this preliminary observation of the candidates' personalities "in action." The tests provided an approach complementary to that of the interviewers. The psychiatrist, for example, could link up the abnormal adaptations of candidates to one another with the adaptations to family, school and occupational environment traced out in his life-history interview. The series included such tasks as moving a heavy object over a series of obstacles, or erecting some apparatus, or having a discussion or debate on some topic, or planning an administrative scheme—say a system of training for future recruits. The situations were so arranged as to display the formation of initial social contacts between members, their behaviour when co-operating on a common task, their reactions to internal disruptive forces (competition between members) and to external forces (competition of one group with another).

Later, other tasks designed to reproduce such officer rôles as independent command, as instructor, or administrator, and as subordinate to senior officers, were posed by the M.T.O. In selecting specialised personnel such as applicants for regular Army commissions, colonial administrators, university short course candidates, and the like, other appropriate "work-sample" situations were devised. Sometimes a task extended over three hours, the group having to study certain material, plan its course of action, and report.

In the early boards, each member studied the candidates independently until the final conference. This led to a somewhat unnatural atmosphere, since members felt constrained not to discuss the current group even informally, and serious disagreements sometimes occurred at the conference, when it was too late to make any further investigation of the doubtful candidates. Thus the introduction of collaboration and mutual consultation at all stages was found to have considerable advantages. Often most of the members would recognise certain candidates as clear passes or fails quite early on, and so feel free to concentrate their study on the borderline or controversial cases. The system had, however, certain drawbacks from the scientific angle, since it became impossible to investigate the reliability or validity of the contributions of the various parts of the board procedure.

Miscellaneous W.O.S.B. Activities

Another function primarily carried out by M.T.O.s was the important one of preparing job analyses, covering the life and duties of each of the main types of officers to be selected. Later in the war, R.T.C. and the boards undertook numerous more specialised jobs including:

(1) Selection of adolescents for eventual training as officers after periods of study at universities or technical colleges.

(2) Diversion of promising candidates who appeared too immature for commissions to special courses of training designed to develop "leadership."

(3) Selection of experienced personnel for special duties—Palestine Mobile Police Force, Civil Affairs officers for administration in the Far East, etc.; also classification of men engaged in the psychological warfare branch.

(4) Guidance of certain classes of officers into suitable employment, including psychiatric disability cases, officers who were surplus in their old jobs, officers who had received adverse reports, and—in particular—repatriated prisoners of war. The latter were dealt with in Officer Resettlement Units, where tests were dispensed with and everything was planned to reassure the men and to readjust them to normal life. Similar principles were extended to the handling of

other rank ex-prisoners in Civil Resettlement Units, and to all demobilised soldiers who felt the need of help in their readjustment. Accounts of this extensive experiment in psychotherapy and its successful results have been given by Wilson (1946) and Curle and Trist (1947).

(5) Selection of war-time temporary officers for permanent commissions in the regular Army, and the development of a technique for selecting temporary officers in the peace-time conscript Army.

(6) Assistance was given also in the selection of Royal Marine Corps officers, paratroop officers, and in an experiment in the organisation and methods division of the Treasury.

Since the war many of the R.T.C. staff have joined together in the Tavistock Institute of Human Relations, and are applying the principles and techniques evolved in the Boards to the study of industrial, educational and other civilian problems (cf. Jaques, *et al.*, 1947).

Civil Service Selection Boards (C.S.S.B.s) have also been set up to advise the Civil Service Commission on the choice of candidates for the administrative class and foreign service. But these differ markedly from W.O.S.B.s. Their chairman and observers have somewhat analogous functions to those of the president and M.T.O.s, but psychiatrists are not employed and the psychological staff consist of ex-industrial psychologists. Group discussions, planning problems and the like are applied, though with the object of throwing light more on the quality or calibre of the candidates' intellectual powers than on their social adjustments.

Discussion

The reader may conclude that W.O.S.B. methods were essentially similar to those of German military psychologists, which we condemned in Chapter I. But in Britain no use was made of elaborate reaction time and other tests alleged to reveal a whole series of faculties or powers of the mind and character. Moreover, much more regard was paid here to the need for validation and follow up, even though the results (summarised in Chapter VII) were not very gratifying owing to the extremely difficult conditions under which they were conducted. It is the M.T.O. side of W.O.S.B. procedure which most closely resembled the German counterpart, and which is probably most open to criticism. None of the leaderless groups or other M.T.O. techniques were standardised tests; no measurements were taken, and only to a limited extent were observations recorded under any standard scheme. Although these techniques aimed to get away from what Gestalt psychologists such as Brown (1936) call "Class Theory", i.e. the analysis of personality into a number of self-sustaining traits, it is likely that most of the psychologically untrained observers continued to treat them as tests of leadership, co-operativeness, and the like. Again, field theory is as yet largely *a priori* and empirically unverified even though it is not—like German theories of character—unverifiable. There is no evidence that the consistency of "reactions to group tensions" is any greater than the consistency of conventional personality traits. We know, for example, that a candidate who displays "leadership" in an artificial test situation will not always show the same trait in real life, but it is doubtful whether his performance in leaderless group tests will be any more prognostic. Indeed, such tests are open to serious criticism from field theory itself. Every group of candidates constitutes a unique combination or Gestalt, hence any one candidate's behaviour is likely to vary considerably with the make-up of the group. Though faced with objectively the same task, his reactions would probably differ if he was tested along with a different set of candidates. This objection was, however, partially met by the common practice of re-grouping candidates in later series of M.T.O. tests. As there were always several sets running simultaneously at a W.O.S.B., the best one or two from each set might be put together, or all the borderline candidates, and so on.

Another fundamental difficulty is that the intense desire of candidates to pass leads them to "put up a show" which may not be typical of their everyday behaviour. All sorts of distorted accounts of what the boards were looking for used to circulate among potential candidates, and they would naturally tend to mould themselves accordingly during the two or three days examination. It is known, for example, that many arrived with ready-made stories to fit the thematic apperception pictures, based on accounts gleaned from previous candidates.

In justification for some of these weaknesses, it should be pointed out that the W.O.S.B. system was not developed, and should not be regarded, purely as a diagnostic technique. Under the circumstances of the war, the fact that it won wide acceptance from the Army and stimulated the recruitment of candidates—in other words its psychotherapeutic effect—was at least as important. It was extremely valuable in that the Army was led to believe that it was thereby getting the best possible officers, even though its methodological shortcomings were such that it probably made numerous incorrect choices. But other selection techniques too are far from perfect, and, like other techniques, it was certainly an improvement on older methods.

Since the war there has been a growing tendency for the selection of managerial, professional and other high-grade workers to be modelled, to some extent, on the W.O.S.B. plan (cf. Fraser, 1947; Hoovers, 1948). One to three day meetings are held instead of half-hour interviews; psychological tests, group discussions and exercises are introduced, and a qualified psychologist usually assists the employers' selection committee. This has several advantages:

(1) It provides a much larger sample of the candidates' behaviour on which to base judgments.
(2) It enables the selectors to observe the social interplay of candidates collaborating or competing with one another.
(3) It appears *manifestly* fair to candidates and to employers, and so stimulates confidence in the system, whether or not this system is entirely sound scientifically.

But a good deal of caution is needed in adopting the W.O.S.B. methods since, as shown later, their value has not been sufficiently thoroughly demonstrated. It should not be supposed that the introduction of a few situational tests, apparently analogous to situations involved in the job, provides the key to accurate assessment of personality and job suitability, nor that the fallibility of human judgment and the need for scientific validation are lessened thereby.

CHAPTER V

SELECTION IN THE ROYAL AIR FORCE

Abstract.—Although aviation psychology had been started in most Air Forces during the 1914–18 war, research did not continue after the Armistice, with the result that in 1939 the R.A.F. was almost entirely dependent on the unstandardised interview in its chief selection situations.

In 1940 a few psychometric tests were introduced under the auspices of the Air Ministry Medical Branch to aid the Aviation Candidates Selection Boards. This testing procedure was under the technical supervision of the Cambridge University Psychology Department where the first groups of W.A.A.F. testers (clerks personnel selection) received their training.

In 1941 Ground Trade Selection tests (including G.V.K.) were devised and introduced by Stephenson (Oxford University) who for two years acted as personnel adviser to the Central Trade Test Board.

Finally, in 1943, the technical responsibility for all selection test procedures became centralised at Air Ministry, where Bott (Toronto University) had founded a training research branch late in 1941. The terms of reference of this branch (now known as Science 4) made it responsible for advice on all psychological matters arising in the Air Force outside the realms of physiology and psychiatry.

The early energies of Training Research were very largely devoted to the reduction of wastage in air crew (especially pilot) training. In 1942 grading (standard flight testing) was introduced; it resulted almost immediately in a 50 per cent. drop in pilot training wastage. Two years later a large battery of tests was assembled with the purpose of obtaining aptitude indices for each category (pilot, navigator, etc.) in respect of all accepted air crew candidates. Reduction in training programmes since 1944 has made validation difficult, but large samples have been followed up wherever possible.

In recent years research has been made into the personality (as distinct from the skill) requirements of air crew. Technical effort

has also been expended in the selection of the officer and apprentice populations, while at the moment an amplified ground trade test battery is in preparation.

In addition a good deal of work has gone into the problem of training assessment (both ground and air), while two other highly important fields (instructional technique and morale survey) claim most of the residual energies of an all too small branch.

The seeds of aviation psychology were sown in the first World War, Italy (Gemelli) and the United States (Henmon and Thorndike) making the most substantial contributions. But, unfortunately, psychologists were nowhere given a chance to follow up their openings in the years after 1918 with the result that the second conflict found the majority of Air Forces unprepared for sudden large-scale expansion. Thus, in 1939, the R.A.F. was relying almost everywhere on traditional methods to select and place its men. The claims of air crew applicants were vetted by a series of Service boards who so far as interviewing was concerned were unselected, untrained and supplied with no technical aids. Similarly the selection of officers was entrusted to boards with little to guide them beyond the knowledge that all the candidates they saw came with some sort of recommendation from a commanding officer. Ground trade applicants who claimed relevant civilian experience were tested in respect of these claims by expert tradesmen; but these same tradesmen were required, with no special training, to undertake the much more general problem of assessing the potentials of the great majority of recruits who made no claim to any special skill. Such very briefly was the picture of selection in the first autumn of the war, and it has to be admitted that some of the features indicated were slow to change appreciably even after the notion of selection as a science had taken fairly firm root.

The introduction of systematic selection in the Air Force was more gradual and less unified than in either the Army or Navy. Three distinct phases are to be noted:

(i) 1940–2. Air crew selection under the technical guidance of F. C. Bartlett (Cambridge University), introduced and sponsored by the Medical Branch of the Air Ministry.
(ii) Ground trade selection initiated and developed by W. Stephenson (Oxford University) acting as personnel adviser to the Central Trade Test Board, West Drayton.

(iii) December, 1941. Arrival at the Air Ministry of E. A. Bott (Toronto University) to found Training Methods (later Training Research and now Science 4). From this point all work on selection and training research gradually became centralised under a single authority.

The early work can most easily be discussed by considering each of these phases in turn.

The "Bartlett Tests."—Early in 1940 the Cambridge Psychological Laboratory was invited to produce a general intelligence test to be given at the Aviation Candidate Selection Boards. The time allowance (15 minutes) was hardly liberal, but the introduction of any standard test, however brief, must be accounted a step forward. Three parallel forms of a 20-item verbal intelligence test (G.I.T.) were prepared and made available to all A.C.S.B.s in the middle of 1940. The tests were at first administered by orderlies, but it quickly became obvious that specially trained staff were needed, and a trade (clerks personnel selection) was created to meet this requirement. The results of the tests, expressed in 5-letter grades based on a 1–2–4–2–1 division of the scores of applicant populations, were available for the guidance of board presidents, but no cut-off levels were laid down, the presidents being free to use their discretion in interpretation. General advice was, however, given on the scores below which admission to (a) pilot and navigator categories, (b) wireless operator and air gunner categories was deemed problematical.

At the same time as the G.I.T. tests, a 15-minute elementary mathematical test was introduced and used as a very approximate guide to educational attainment. Both the G.I.T. and E.M.T. had the merit of being objectively scored, the scores being evaluated against standard grades. A third written test (general knowledge) consisted of two fifty-word essays the purpose being to afford some clue to powers of expression and alertness to current affairs. Only a very slight reliance on this test was, however, encouraged, the risk of subjectivity in marking being considerable.

The three tests just described remained part of the standard air crew selection procedure for more than three years. They may not have made a large contribution to the training wastage problem, but they undoubtedly helped to make the Service test-conscious, and in doing this paved the way for the much fuller testing programmes of later years. In the beginning of 1942 the battery was augmented by the S.M.A.3 (pilot co-ordination) test (cf. p. 248), and a sub-battery of three tests—directions, tapping and morse aptitude—aimed at the elimination of unsuitable wireless operator trainees. S.M.A.3 was the work of G. O. Williams (Central Medical Establishment, R.A.F.), while the wireless battery was assembled by G. C. Drew (Cambridge University), who later came to do full time work at the Air Ministry. The latter tests were introduced executively following an experiment at Blackpool which led to a sharp reduction in wastage in the basic signals course.

The "G.V.K." Tests.—In the early part of 1941 Stephenson came from Oxford to take up full-time duties as personnel adviser (ground) to the Central Trade Test Board, with whom he remained for more than two years. Very little experience of trade allocation problems convinced him of the need for a test whose first function would be to sort the large and diverse weekly intakes into some half dozen broad ability zones. For this purpose categorisation on the most general lines was required, the question of group and specific abilities not being of paramount interest at this stage. Because, however, sharply contrasted trades, e.g. clerical and mechanical, were liable to make demands for personnel at each ability level there was a clear advantage in designing a two-purpose intelligence test which would give pointers in respect both of general ability and of strong leanings towards practical or theoretical work. The three-part test known as G.V.K. is well designed to do both these things. A reliable estimate of a recruit's general level can be obtained by averaging the sum of his three scores. But as the "V" section has a considerable language (as well as verbal intelligence) content it serves as a rough guide to educational attainment and suitability for sedentary occupation, while "K" directs attention to the possession or non-possession of practical-mindedness. A simple arithmetic test was added to this short battery which in all occupied rather less than an hour of testing time.

To understand how the above tests were used it should be explained that during the war years the ground trade problem was one of trade recommendation, not of overall acceptance or rejection. In this it differed sharply from the air crew situation where initial rejection rates rarely sank below 50 per cent. But the R.A.F. ground population was in effect no different from the population applying for entry and the divergences of both from the civilian population were probably small (cf. p. 198).

Recommendations to ground trades were made by interviewers chosen primarily for their experience in certain technical trades. These were the people whose task the test battery was intended to assist. For such aid to be effective, guidance was given on the mean test scores found among trained personnel in each R.A.F. trade, and on a "failure" score for each trade, i.e. a score below which success in training had been demonstrated to be small. (Samples of this are given in a later Chapter.) Scores were presented in percentile form, the basic population being a random sample of 5,000 drawn from early 1942 intakes. The norms established at this date have been held constant ever since, but revision in terms of current N.S.A. and volunteer intakes is intended.

Specialised tests were devised only in respect of one or two priority trades, e.g. radio wireless mechanics. These were given at the interviewer's request to recruits found potentially suitable for such trades. It will doubtless be thought that a battery of four short general tests and two or three special ones is a very slight instrument with which to attempt allocation to more than a hundred occupations. Unquestionably an ampler battery would be desirable and it is hoped to introduce one in the comparatively near future. Meantime one can note the ingenuity and economy with which the G.V.K. test performs its double function.

Training Research.—Both air crew and ground trade selection were thus already under way by December, 1941, when Bott, with C. R. Myers as his assistant, took up his Air Ministry post as adviser to the Air Member for Training. For twenty months an embryonic branch was concerned almost exclusively with air crew problems, particularly with that of drastically reducing pilot training wastage. Its terms of reference were, however, exceedingly wide; so wide, in fact, as to cover almost any application of psychology anywhere in the Air Force. It was thus possible when Stephenson relinquished his post in July, 1943, for ground trade selection to be brought under the same head as the selection of air crew, while the fact that the advisers had the right of entry to both selection and training problems made possible a longitudinal survey of the airman's whole career. That the study of selection and training has proceeded unevenly, the bulk of enquiry in the early years having been directed to the former, must not be ascribed to any maldistribution of effort. It has resulted simply from the fact that it is comparatively wasteful to undertake detailed research into training methods until something has been done to ensure that approximately the right people are entering training.

Between 1942 and 1945 Training Research (the name was altered from Training Methods early in 1943) expanded from a section with three or four research workers to a branch with a provisional scientific establishment of seventeen. During most of this period technical staff were recruited by a series of special appointments, but in the autumn of 1944, it was decided to bring these temporary appointments within the framework of the Scientific Civil Service. During a large part of 1945 nearly all the established posts were filled, though lack of civilian psychologists led to a number being taken by serving officers with psychological qualifications. Since the end of the war, however, the strength of the branch has declined by over 50 per cent., the chief reasons being demobilisation and delays in deciding the future status of psychologists in Government employment. At the moment of writing (December, 1947) the small staff remaining is being forced to spread its energies among a greater number of problems than it can possibly do full justice to, while other problems, some of great urgency, cannot be tackled at all.

From July, 1943, when A. J. Marshall took over Stephenson's work on ground trade selection, the branch was divided into two sections, the first dealing with all air crew matters (whether selection or training) and the second with ground trade problems. Later re-organisation brought all selection problems (air and ground) together, a second section being set up to deal with training problems.

The functions of the branch have throughout been advisory. In the early days responsibility was to the Air Member for Training alone; later, as the emphasis shifted back to the earliest stages of selection and as the desirability of a centralised selection system came to be appreciated throughout the Air Ministry, advice was increasingly sought by the Air Member for Personnel also*.

---

* In 1947 a Deputy Directorate of Selection was formally established to control the policy and administration of all selection procedures throughout the Service. This personnel branch had, in previous years, done much to raise the general standard of service interviewing, particularly in the fields of reselection of air crew training failures, reallocation of tour-expired air crew and the disposal of former prisoners of war.

Finally, as Science 4, the branch became part of the Scientific Adviser's organisation and through him responsible to the Chief of Air Staff.

Training of Testing Staff

A short account of the selection and training of testing staff may be of interest. From 1942 to 1946 Clerks P.S. consisted exclusively of women, but the quickened release of W.A.A.F. brought about an almost complete change-over to R.A.F. personnel. It is anticipated that the final constitution of the trade will provide for equal proportions of men and women. As regards selection two main points have been established. First, it has been found unprofitable to train anybody below the top quarter of the R.A.F./W.A.A.F. population in general intelligence. Secondly, great pains have to be taken to ensure that every volunteer knows what the duties of the trade involve. The second point may seem obvious, but very little experience showed that the notions of many early applicants as to the nature of psychological work were wild and sometimes crude. A number assumed, rather naturally, that personnel selection duties entailed a great deal of individual interviewing, while others, less creditably, looked forward to a banquet of Freudian symbolism.

Pre-selected personnel received their training in courses originally of three weeks duration, but the length of these courses has grown to six weeks, as the scope of the work has developed. Not only has the number and variety of tests and testing situations in the R.A.F. increased steadily since 1942, but features other than test administration have been added. Thus while the prime aim of the courses has always been to bring testers to a high level of proficiency in giving and invigilating tests under standard conditions, it has also been found profitable to give a grounding in statistics, test construction and research planning. As a result of this the limited resources of the scientific staff have been saved from a mass of semi-technical effort which could never have been entrusted to untrained assistants. In addition the existence of a corps of intelligent and enthusiastic testers has been of inestimable value in spreading an enlightened conception of psychological aims among the Service.

A great deal of thought has gone to the question of course content and syllabus structure. Instructional periods are divided almost equally into practical (i.e. exercises in the giving of patters, demonstrations by wire recorder, marking, practice, etc.) and theoretical (mainly but not entirely lectures). Lectures include three main subjects, principles of mental measurement, statistics (where each lecture is followed by an hour of practice) and R.A.F. selection situations. About three-quarters of all the instruction is given by the school staff (personnel selection officers), the remainder being covered by visiting lecturers from Science 4. During the last week a passing-out examination takes place, the main emphasis being on practical testing with objective tests on testing theory and statistics.

Air Crew Selection

The entire choice of air crew (medical issues apart) was entrusted at the beginning of the war to boards of serving officers who were afforded no more definite briefing than to find the right types. Three assumptions can be detected behind this: (a) that there are right and wrong types for air crew duties, (b) that such types can be detected by simple interview, (c) that officers who have themselves been air crew are particularly suited to do the detecting.

The questionableness of these assumptions needs no elaboration; more important is it to appreciate the complexity of what the boards were asked to do. First, they were expected to decide not merely who should be accepted as air crew and who should not, but also to which air crew category an accepted recruit should be sent for training. Secondly, the boards were given no guidance as to the relative importance of personality and skill factors in reaching these decisions, and thirdly, they were provided with no technical aids with which to measure either. The introduction of the Bartlett tests marked a small but significant step towards the solution of the last issue. Whatever intelligence, educational attainment and sensori-motor co-ordination may be, they clearly fall outside what are ordinarily considered personal qualities. This distinction once made, the path was to some extent cleared for the separation of skill assessment from the appraisal of character. This separation was carried much further by the introduction of the pilot "grading" test in 1942, and was completed when the Air Crew Aptitude Test Battery was brought in during the spring of 1944. From that point the job of the selection boards was to evaluate personal suitability for air crew, the allocation to categories being undertaken at a later stage and being determined by aptitude and preference*.

Towards the end of 1941 the decision to move flying training overseas focused attention on the selection of pilots. This was the first concrete problem Training Research were called on to attack and their initial analysis showed some alarming figures. It was found that of many hundreds who had commenced pilot training two years earlier only 41 per cent. had entered on operations, while 36 per cent. had failed to qualify. Moreover, two-thirds of these failures were falling short at the elementary flying stage which, in other words, was acting as a selection screen so far as flying ability was concerned. This was uneconomical enough when training was in this country, but it would clearly be insupportable with schools spread over half the globe.

The problem was far too urgent to allow of the evolution of a test battery even had such a battery seemed the most likely way to reduce wastage. It is, however, extremely questionable whether such an approach could ever provide as good an answer as the work sample method actually adopted. This method entailed a short period (12 hours) of flying training with standard tests at prescribed intervals (originally at 7 and 11 hours, later at 5½, 8½ and 11½ hours).

It will be seen that such a selection system assumes a close correlation between speed of learning and the ability subsequently shown. This assumption was firmly substantiated before the flight tests were brought in, and the whole justification for their use is based upon it. The evidence is given in Chapter XVI, together with details of test content, weighting of items, scoring and validation, and other technical points arising from the need to operate a very complicated type of test in a number of flying schools simultaneously (cf. also Parry, 1947). As the word grading implies, the job of each of these schools was not to pass or fail the individuals it tested, but to array them in order of merit so that the cream of each intake could be skimmed by a central authority according to the needs of the moment. To utilise the tests to the full it was obviously necessary to grade considerably more than would be required to enter training, but, this margin conceded, the scheme was as well suited to picking the best 20 per cent. as the best 60 per cent.

---

* This statement slightly exaggerates the cleavage. If intelligence and educational attainment are regarded as aptitudes then aptitude did play some small part in the A.C.S.B. procedure. It is, however, true that temperament and character assessment played no part at all in category classification, nor is any evidence to hand to suggest that they should have done.

Early in 1942 gross pilot training wastage had been approximately 48 per cent.; with the introduction of grading the figure dropped to nearly half (25 per cent.) The most important reduction was very naturally at the elementary stage where the gross comparative percentages were 30 and 14, but a perceptible shrinkage (11 to 9) resulted at the next stage (service flying), while operational training wastage fell from 7 per cent. to 2 per cent.

The introduction of grading inevitably affected the functioning of A.C.S.B.s which hitherto had decided the air crew categories each accepted applicant should enter. It was part and parcel of the new system that a high proportion of those sent to grading should not become pilots, and consequently re-classification of these to some other category had to be provided for. To meet this the boards early in 1942 were instructed to classify to the pilot, navigator and bomber categories jointly instead of severally. This meant that all deemed suitable for any one of these categories were given the opportunity to go to grading, and as about 90 per cent. of applicants were anxious to do this the field for the final selection of pilot trainees was greatly increased. The scheme, however, did nothing to aid the selection of the other categories and the need for a further selection phase which would permit of the assessment of navigator, gunner, etc. skills, became apparent eighteen months later.

The primary purpose of aptitude testing was to secure skill measures for air crew accepted by the selection boards in respect of each of the six main categories. This led to the preparation of a two-day testing programme with twenty-four tests yielding six sub-batteries. No experimental period was available either to devise new tests or validate tests loaned by other services*. *A posteriori* validation was, of course, planned, and has been carried out so far as continuity of training has allowed—a summary of the findings appears in Chapter XVI, together with a detailed account of the way the problem was handled. Here it need only be added that weight was given in classification to the recruits' preferences wherever skill as measured by the tests was adequate and the needs of the Service permitted acceptance.

---

* Tribute must be paid to the U.S.A.A.F. in particular, who loaned a considerable number of their paper tests.

From April, 1944 then, scientific measures have been in use for the assessment of air crew skills. But the problem of personality assessment remained, and though this has been attacked during the last two years much still remains to be done. A job analysis based on the opinions of 350 air crew officers has at least shown close agreement about the personal qualities held important in air crew, and incidentally suggests that the needs are much the same from category to category. It has also shown that the important qualities are for the most part conspicuously hard to rate at interview, while those that are easily assessable (diction, appearance and the like) are reckoned of very little account. All this makes it a difficult matter to supersede the unstandardised service interview by a more reliable instrument. But the old procedure has been strengthened by the insertion of a pre-interview at which the qualities deemed important receive independent rating. Table XLV (p. 277) shows that the reliability of trait assessments is on the whole satisfactory. In addition a biographical inventory is being prepared which is expected to throw more light on this question.

Though the separation of skill and personality assessment has been a logical development, their temporal and geographical displacement has been an accident of circumstances. A fully developed selection system works not by a series of successive hurdles, but provides for a weighing one against the other of at least the most important factors. The R.A.F. have now taken the first step towards this by placing air crew selection boards and aptitude testing at the same centre. The next technical problem is to find a reliable way of evaluating personality assessment so that this can be related to aptitude scores.

To complete the picture of air crew selection two further features should be touched on:

1. Advantage is now taken of the first stage of all—combined recruiting centre—to eliminate the more hopeless applicants. This is done by means of a fully controlled interview designed to anticipate selection board verdicts. It has been found possible to anticipate in respect of the worst 5 per cent. to 10 per cent. with almost 100 per cent. accuracy. Technically this is purely a screening interview, and is of interest because it has a precise and limited objective and has been proved to work effectively in the hands of interviewers with very little training.
2. It has been claimed by the R.C.A.F. that the Link trainer

can be adapted as a selection test to yield results comparable in accuracy to grading. This, if true, would result in a great economy of money, time and effort. An experiment to test the claim was planned in April, 1945, but large scale cessation of flying training has resulted in a mere trickle of validatory evidence. The few cases to hand suggest that Link adds appreciably to aptitude testing; but it has not yet been confirmed that it adds as much as grading. There is also a very real danger of candidates getting access to a Link before coming up for selection, and thereby securing a great advantage.

It will perhaps be well to summarise this rather complex section with a table showing the successive air crew selection phases:*

Name of Phase	*Function*	*Selection Instrument*	*Action*
1. Combined Recruiting Centre	Loose Personality Screen	Controlled Interview	Elimination of lowest 10-25%
2. Combined Selection Centre	A. Full Personality Assessment	Pre-interview Technique	Results of A and B weighed by Final Board with power to interview further
	B. Differential skill Assessments	2-Day Testing Programme	
(A and B prior to mid-1947 constituted successive hurdles)			
3. Grading. (Potential Pilots only)	Assessment of skill in the air	Standard Flight Tests after 5½, 8½ and 11½ hours	Better pupils to Pilot training. Remainder to other categories found suitable in 2B

Selection of Aircraft Apprentices, Administrative Apprentices and Boy Entrants

The main selection situations in the R.A.F., outside air crew, are ground trades, apprentices and boy entrants, and officers. In the ground trades the procedure instituted by Stephenson (p. 70) has so far received no fundamental re-direction, though additions and modifications have been made to it. Further information as to its working and its effectiveness is supplied in Chapter XVI.

In 1939 apprentices and boy entrants were selected exclusively on educational attainment. Important though this factor is it cannot be relied on to yield consistent indices of technical ability, and when these entries were re-opened after the war the need for an amplified selection procedure was generally recognised. The current programme for aircraft apprentices is in two phases. The first,* carried out in a number of local centres, consists of a qualifying educational examination as a result of which a small proportion (under 10 per cent.) of applicants are eliminated. At the same time a general intelligence test is taken. All who survive this stage are then called to the R.A.F. selection centre where they undergo medical examination, take a one-and-a-half day programme of aptitude tests,† and are given a two-stage interview. Their final acceptance or rejection is determined by three lines of evidence—education, intelligence and technical aptitude, brought together and interpreted by a trained interviewer. The tests are additionally used to give pointers to trade allocation, particularly as between the mechanical and electrical trades, although every effort is made to evaluate the strength of trade preference and to concede first choices when possible. The tests in use have been selected from a large battery administered to the first post-war intake and provisionally validated against early training results. Full validation will be spread over a number of intakes and will take all stages of the four-year training into account.

---

* Grading has now been discontinued for administrative reasons, pilot skill being assessed by aptitude tests alone.

There is no preliminary qualifying examination for administrative apprentices and boy entrants. The first group are called to the selection centre where they take tests in educational attainment, general intelligence and clerical aptitude. A level of general ability at least as high as that for aircraft apprentices is required.

The two apprentice populations constitute the technical nucleus of the R.A.F. and between them provide a large proportion of the personnel in the high-grade trades. The information arrived at in developing their respective test batteries is consequently directly relevant to the wider task of building up a general ground trade test battery. Further light is expected from the selection of boy entrants who are required for feeder trades.

Administrative apprentices and boy entrants are, like aircraft apprentices, interviewed during the selection stage, their trade choices being carefully explored and effective contact made with every applicant.

---

* This stage of selection is handled by the Directorate of Educational Services.

† A considerable debt is owed to the N.I.I.P. for their generous advice and loan of tests.

Throughout the war the R.A.F. relied on the interview for its choice of officers, but in 1945 it was decided to adopt W.O.S.B. methods. A small preliminary experiment (December, 1943) in which R.A.F. candidates for commissions were put through both selection procedures had pointed towards the new technique as the better predictor of O.C.T.U. results. In June, 1946, an R.A.F. Selection Board with Air Commodore as president was established to choose the first post-war entry of cadets for the R.A.F. College, Cranwell. Board personnel had the advantage of observation and training at Army centres.

During the following autumn the same board was called on to evaluate the claims of serving officers who had applied for permanent commissions, and since then it has been in almost continual session, handling in turn both types of entry.

It is early to look for validation evidence in respect of either the cadet entry or the serving officer population. Indeed, in view of the high rejection rate it is likely to be extremely hard to evaluate follow-up results, particularly in the case of the cadets where the great proportion of failures are lost to the Service completely.

Other Aspects of Air Force Psychology

The main types of problems, outside selection, with which Science 4 has been concerned are training assessments, instructional method, and morale. The first of these has received a fair share of attention, the second has been studied only in respect of isolated problems, while the third is claiming consideration now for the first time.

Training Assessment.—The general problem is the introduction of objective methods in respect of practical as well as written tasks. The unreliability of traditional methods has been demonstrated over and over again and hardly needs emphasis here. Two illustrations will suffice: Variations in the use of a common scale have been shown to be so great from one school to another that the pooling of marks from different schools often leads to the most grotesque results (this, of course, is compatible with objective orders of merit within each school). Secondly, scarcely any agreement has been found, in many instances, between the marks obtained by individuals at similarly named subjects taught in successive stages of training.

The objectification of written examinations presents no special difficulty in R.A.F. technical trades. The best procedure involves the summoning of a small panel of instructors who under the guidance of a psychologist elaborate a pool of new-type items. From this it is usually possible to extract two or three alternative versions each with eighty to a hundred questions.

The difficulties with practical tasks and above all aerial exercises, e.g. pilot skill, are very much greater and a long experimental period is customary before even moderately trustworthy tests can be attained. The phase check or standard efficiency test (p. 267) which entails the application of the objective method to learned drill sequences has been adopted in many settings with satisfactory results.

Two by-products of the whole training assessment problem are the supplying of reliable criteria against which to validate selection tests and the application of up-to-date corrections to syllabus content. For the pooling of objective items forces attention on the emphasis given to different areas and in some cases on the survival of features which no longer need be taught. The danger of accepting a neatly-made assessment scheme as evidence of sound assessing has been demonstrated again and again, as also the danger of assuming that instruction and assessment should necessarily be undertaken by the same people.

Instructional Methods.—Apart from a few isolated experiments (mostly in the field of morse training) very little has been undertaken here. This is not for any lack of problems. On the contrary the teaching syllabuses of most technical trades have become so swollen and overloaded that there is a crying need for basic investigation. What is required ideally is the attachment of a qualified psychologist to each of the main R.A.F. schools. Something much more fundamental than piecemeal patching and shuffling is called for. Controlled experiments into such things as length of working day, relation of theory and practice, training sequence, classroom conditions, use of demonstration aids, use of notes and selection of instructors all need to be investigated.

Morale.—Peace-time morale, particularly with the conscription issue to reckon with, is a far more difficult matter than morale in war. The first step is the systematic evaluation of opinions on such topics as general sources of discontent, accident prevention and attitude of the conscript to the Service. An attempt is now being made in these directions which make their appeal very naturally to the social psychologist rather than the psychometrist. While evaluation is in progress the Service psychologist can do much by making the Service mind aware that these problems exist and have to be met. It is usual and no doubt natural that these efforts should often be met with resistance. The view that people will always grouse, or that the conscript is bound to resent his period of service, is frequently heard, and needs to be combated.

PART II

The principles of personnel selection and guidance which have evolved both from pre-war investigations and from war-time experience.

CHAPTER VI

GENERAL PRINCIPLES OF VOCATIONAL CLASSIFICATION

Abstract.—Personnel work in the Services consisted of a combination of selection—in the sense of picking the best men for a job, and of guidance—in the sense of recommending the job which would best suit each man. This conception, which we call vocational classification, is worth pursuing in peace-time education and industry. The following are the main elements of classification procedures:

(1) Provision of information about jobs to candidates helps to promote self-guidance. The staff responsible for classification must also be provided with information in a usable form, but do not usually need detailed job analyses.
(2) Collection of relevant data about adult candidates is mainly done by biographical questionnaire and interview. But cumulative pupil record cards are more appropriate with children.
(3) Educational and vocational classification should take account not merely of the candidates' abilities, interests and personalities, but also of sociological factors, in particular, attitudes towards education or to employment.
(4) Tests and measurements are important, but their rôle in classification should not be exaggerated.
(5) Interviewing has several vital functions and, in spite of its weaknesses, cannot be dispensed with.
(6) Classification should not be based on a single "cross-section" of the individual, but should be linked as a continuous process with education or training.
(7) The main feature of the scientific approach is validation, or continual checking of the value of the procedures employed.
(8) Fully qualified psychologists can seldom undertake classification themselves, but should give thorough training to suitable persons in education, industry, etc., and should supervise their work.

The work of psychologists in the Forces which we have been describing can hardly be called occupational *selection* or vocational *guidance*. It is a blend of both, which we will call vocational *classification*. The term is not an ideal one since it suggests the forcing of recruits into jobs which they may not approve; whereas the essential object is to adjust or harmonise the whole range of recruits (each possessing his or her individual capacities and interests) with the range of Service employments (each having its particular requirements).

Occupational selection has in the past often been a somewhat anarchical procedure whereby an industry attempted to grab the best available workers, and to disregard the wider interests of the community. It will be shown later that, provided the supply of men for a job greatly exceeds the demand, the quality of employees can be raised by almost any selection procedure, whether it be an interview by an untrained foreman or a few tests whose validity is very low. Naturally the better the procedure the greater the improvement in quality, but even an inefficient procedure will appreciably cut down the number of unsatisfactory employees so long as no account is taken of the candidates wrongly rejected (cf. p. 127). If, however, we pay attention not only to the candidates wrongly accepted, i.e. those who pass the selection procedure and turn out unsatisfactorily, but also those rejected by the procedure who could actually have managed the job, the procedure must attain a much higher validity. Again, if the supply does not greatly exceed the demand, and the so-called selection ratio (the number to be employed divided by the number of candidates) is high, it is far more difficult to choose the best men for the job (cf. Tiffin, 1946).

In 1941, when psychologists entered the Forces, anarchic conceptions were rife. Certain branches of the Navy, Army, and Air Force were using their own interview or other procedures to grab high-quality recruits, entirely disregarding both the needs of other less powerful or less attractive branches, and the interests of the thousands of men they rejected who would have been perfectly adequate for the work. Some branches, as we have seen, even enlisted the aid of university psychologists to improve their procedures. The results were quite gratifying to the branch concerned and to the psychologists, but actually produced greater vocational maladjustment in the Forces as a whole. The same conception was prevalent in education until the 1944 Act. "Public" and Grammar Schools could afford to set stiff entrance and scholarship examinations, whose validity in picking the pupils best fitted for secondary education was dubious, because of the low selection ratio, i.e. the excess of candidates over available places. It did not matter to them that thousands of pupils were rejected who would have derived as much (or greater) benefit from advanced education as the acceptances.

Vocational guidance, on the other hand, has tended to regard the interests and capabilities of the individual as paramount, and has recommended the job or types of job best suited to him. While N.I.I.P. psychologists have always taken into account the opportunities that each individual would have for following the recommended career, they were naturally unable to match their recommendations closely with the state of the labour market. Their system, in other words, works best when demand exceeds supply, so that the individual can readily pick and choose among a wide range of available jobs; whereas vocational selection prefers the opposite state of affairs. Other difficulties with the guidance given by the National Institute to individual applicants are its length and expense. When each candidate consumes the best part of a day's work by a highly trained investigator, the fee cannot be reduced much below £4, which puts it out of reach of the majority of the population. Such guidance, moreover, is largely restricted to the higher occupational levels such as the professions, and a much simpler procedure might be adequate for less skilled work in which individual aptitude and interest in the work itself are less important.

Very early then in the history of the Institute, attempts were made to extend the procedures more widely. In a series of experiments in London, Birmingham and elsewhere, guidance was given to large groups of school-leavers who were representative of the general population. Outstanding features of this work were its comparative speed and simplicity, its application to all occupational grades, its validation by subsequent follow-up, and its integration with a survey of the labour market in the particular districts where it was carried out. Thus, in the most recent Birmingham experiment, some 1,600 pupils were psychologically assessed by 74 specially trained teachers, and were placed in jobs which accorded as closely as possible with the assessments, by 10 juvenile employment officers (cf. Hunt and Smith, 1945). Rodger

(1939a) suggests that large-scale vocational classification of school-leavers can best be tackled by such co-operation between the schools and the Employment Exchanges. Similar schemes have been attempted elsewhere, for example, by Macdonald (1938) in Edinburgh and Meiklejohn (1945) in Ayrshire.

In the Forces, as we have seen, it was essential to find suitable employment for all recruits and to consider the requirements of all the branches. The supply of good men usually fell far short of the demands, and wrong rejections by inefficient selection procedures had as serious effects as wrong acceptances. Vocational guidance, in the sense of recommending for each recruit the job he could do best, was equally inappropriate, since only a limited range of employments was available, each requiring certain definite quotas of men or women. Thus it was impossible to suit the interests, experience and aptitudes of every individual. To take a naval example: in 1942–3 most of the mechanic branches were crying out for men, and large numbers of recruits with no engineering experience had to be trained for such jobs. But by 1945 the demand dropped so greatly that many recruits with good mechanical ability and training had to become seamen, writers (clerks), and so forth. Even if Admiralty psychologists had possessed a perfect test which showed that, say, 20 per cent. of recruits had good mechanical ability, they would have had to disregard its findings when the Navy wanted either 30 per cent. or 10 per cent. of recruits for mechanic branches. While fluctuations may be less violent in civilian life, here too there is no guarantee that the distribution of abilities and interests of school-leavers in any one area will match the distribution of job requirements. And even in the largest industrial concern the range of jobs open to employees about to be allocated or transferred is obviously still more restricted. The present lack of candidates for mining, cotton, brick-making and other industries shows how impossible it is to obtain correspondence. A good illustration was provided by the Dundee vocational guidance investigation (Pallister, 1938), when it was found that roughly 16 per cent. of the girls leaving school would be absorbed by each of the main employments—factory work, jute mill, shop assistant, and office work. But when a questionnaire was given to typical pupils, 30 per cent. wanted shop assistant, 27 per cent. factory, 20 per cent. office, and 4 per cent. mill. Clearly a large proportion of girls were certain to be disappointed.

The conception of classification worked out in the Forces has important civilian applications during the post-war years when Britain's economic recovery largely depends on the efficient utilisation of the available man- and woman-power. It is implicit, too, in the new approach to education. The desire to provide the amount and the kind of education that accords with each child's talents and interests is praiseworthy, even though, as Burt (1943a) shows, the notion of classifying pupils at the age of 11–12 into academic, technical and modern "types" is psychologically unsound. As mentioned in Chapter I, most American high schools and colleges have educational counsellors who are trained in the application of modern psychological techniques of testing and interviewing, and who are likely, therefore, to be able to guide the educational careers of the pupils and students more effectively than, say, an over-worked head master. In industry also the conception of classification is winning wider recognition. A firm such as Rowntrees at York absorbs every year a certain proportion of the school-leaving population of the district, and uses its personnel selection procedures not so much to pick the cream as to sort the available employees according to the various openings within the factory for which they will be most suited (cf. Northcott, 1945). Tiffin (1946), an American industrial psychologist, similarly points out that scientific vocational methods do not lose their value when labour is in short supply, since it is all the more important to use them for transfers and promotions within any one firm, in order to improve the quality of employees assigned to each job, and to improve morale by reducing the number of misfits.

What are the main principles underlying large-scale classification procedures as worked out in the Forces, which can also be applied in modern industry and education? They may be listed as follows:

(1) Provision of information to candidates and to those responsible for the classification.
(2) Collection of relevant data about the candidates by questionnaires, cumulative records, etc.
(3) Due consideration of sociological factors.
(4) Application of standardised tests of aptitudes and abilities.
(5) Competent interviewing.
(6) Linking of the selection or guidance process with education and training.

(7) Validation and follow-up of all procedures.
(8) Training and supervision of personnel who carry out the scheme.

Provision of Information

Macrae (1932), Oakley (1937) and others have pointed out that one of the most potent sources of vocational maladjustment is ignorance or distorted information as to the nature of various types of job. Experience in the Navy in particular shows that most people are capable of a large measure of self-guidance provided that they have access to simple but accurate descriptions of the work and all that it involves; and that a large amount of the psychologist's time can be saved if they have received and considered this information before seeing him. While direct vocational education is seldom if ever advocated nowadays, there is everything to be said for more education about vocations. In enlightened primary schools projects are often devised, which lead the pupils to acquire English and arithmetical skills, and these may be centred round, say, the post office, the railways, ships, house-building, or a local industry. As the leaving age draws nearer, more comprehensive surveys of employments can be provided by films, books and pamphlets, visits to works, and, as Rodger (1944) suggests, "Brains Trust" discussions. Talks by employers are, of course, likely to over-emphasise the attractions of certain jobs, but talks by young or by mature employees, e.g. former pupils, may be enlightening. The value of giving vocational information to beginners in any large industry, by films, talks, and works tours, is widely recognised, since it not only helps in the initial process of allocation, or in the re-allocation of employees who prove unsuitable, but also stimulates each worker's interest in his particular job when he realises how it fits into the whole scheme of production.

Still more necessary is it for the person responsible for guidance or selection to be familiar with job requirements, and to be able to clear up misconceptions in the minds of candidates. Teachers who assess and advise school-leavers cannot be expected to be industrial experts, though any knowledge they can acquire and use, or pass on, is better than none. They should for example, study Oakley's Handbook (1937), the Ministry of Labour's pamphlets on careers, and other such literature. But, just as in the Services, so in civilian life the state of the labour market varies, and the nature of many jobs may alter radically. Hence it should be the Ministry of Labour's function to survey the openings and to give adequate up-to-date information to the schools. Rodger (1939a) suggests that this should be done under the same headings as the teacher or psychologist uses in assessing the qualifications of the candidates, for example:

(1) Physical and medical requirements.
(2) Educational qualifications.
(3) Level of general intelligence.
(4) Special aptitudes.
(5) Interests.
(6) Disposition or temperamental qualities.
(7) Other relevant circumstances (training, apprenticeship, etc.).

It is not possible to enlarge here on scientific job analysis procedures (cf. Bingham and Freyd, 1926; Viteles, 1932), but the main desiderata should be mentioned. Neither superficial observation of the work by a psychologist, nor the description given by an employer or foreman are adequate. The experience of men who are successful at the work and of failures should be critically examined and integrated, and special note made of the commoner causes of failure. The analysis should avoid the use of vague traits and abilities (persistence, concentration, dexterity and the like), and should present as objective as possible a description of the actual duties or operations performed. At the same time it is seldom useful to list in detail the specific movements, etc., or to resort to time and motion study analyses. Rodger's list of points can be expanded to include information about employment conditions (physical and social), special advantages and disadvantages, pay and other incentives, promotion and prospects, length and nature of training, relation to other jobs in the organisation, and types of experience that may be useful or necessary.

Collection of Data

N. A. B. Wilson estimates that at least half of the information on which vocational decisions are based should be derived from past history, the other half consisting of tests and interview judgments of present abilities and other qualities. Hence the enquiry into the past must be very carefully planned.

An account is given in Chapter VIII of the biographical questionnaires adopted in the Forces, and of the value of the information which they yielded. On these are collected not only details of the recruits' background, but their test scores and the interviewer's findings and recommendations. Thus they constitute the keystone of the whole classification procedure. They are nevertheless limited both by the accuracy of the recruits' memory, and by the time factor. They cannot contain more questions than the dullest recruit can understand and answer in half to three-quarters of an hour.

Vocational work in schools has the great advantage that it does not have to be carried out in a few hours or days. Relevant features of children's careers can be observed and recorded at the time, instead of having to be recalled later. Thus, the cumulative record card should be made the focus of classification procedures among children. (In an industrial firm, an analogous record card is obviously of the greatest assistance to the personnel official who deals with transfers and promotions.) While most education authorities now maintain pupil record cards, an investigation by the N.I.I.P. (Cockett, 1945) has shown how unsatisfactory are most of these for vocational purposes. Their emphasis tends to be on the weaknesses of pupils, whereas the careers master is mainly concerned with potentialities and strengths. Usually they are so incomplete that he will need to design a supplementary form or else adopt a carefully constructed card such as that of the Institute, which covers:

(1) Records of scholastic attainments over several years.
(2) Medical data.
(3) Positions held in school, games, etc.
(4) Family background, out-of-school employment.
(5) Interests and hobbies.
(6) Ratings of traits of temperament and character.
(7) Objective test results.
(8) Occupational suggestions.

Hamley's (1937) and Fleming's (1945) suggestions are also worth consulting. Such a document is preferably filled in by the careers master or other specially trained member of the staff on the basis of discussions with the children's teachers, rather than by the various teachers themselves.

Sociological Factors

So numerous and so complex have been the developments in psychological and statistical techniques in recent years that there is often a tendency to neglect the wider sociological setting of problems in applied psychology. As Burt (1947) has described, selection for secondary education is often discussed purely in terms of the children's abilities as measured by examinations or teachers' gradings, intelligence or other psychological tests. It is not enough to add assessments of personality traits such as industriousness, or of interests in different types of curricula. In many cases the main factor in success or failure at the secondary school is the attitude towards schooling current in the social group from which the children come, and particularly in their families. Similarly, estimates have been published of the proportion of the population likely to possess the ability needed for university education, apparently ignoring completely the fact that going or not going to a university is largely determined by the attitudes of the social and economic classes to which prospective students belong. When some new technique, say intelligence tests or vocational guidance, works well in one context, over-enthusiastic advocates apply it indiscriminately, not realising that it may require considerable modifications to adapt it to the mentalities of the individuals, or to fit it into the framework of the social institutions with which they are concerned. We have mentioned in earlier chapters some of the conflicts that occurred between the technicians and Service authorities. Undoubtedly these would have been far more serious but for the constant endeavour of the psychologists to adjust their procedures to Navy and Army traditions, customs and standards. Similarly the educational or vocational psychologist should avoid planning any scheme of selection or guidance in the abstract, and should make himself thoroughly acquainted with the educational or industrial system within which he is working. Burt deplores the absence of sociology from the training of most educational psychologists, and points out that the provision of psychiatric social workers in many child guidance clinics is an inadequate substitute, since their job is to note abnormal factors in the home situation rather than the normal mores of the society in which the children are reared.

To the Tavistock Clinic Group (cf. p. 64), it is not merely sociological but "sociatric" factors which are of prime importance, that is the underlying—often unconscious—emotional forces within the organisation. It is well known, for example, that the grievances consciously expressed by employees in an industrial firm, say about the canteen or about wages, are often projections of, or unwitting substitutes for, deeper dissatisfactions, with the autocratic attitudes of the management. Similarly a request from a firm for the institution of scientific selection methods may be a defence reaction, an unconscious desire to ward off enquiries into more serious maladjustments. If the psychologist meets such a request by devising selection tests, or if welfare officers improve the canteen arrangements, these fragmentary solutions may actually exacerbate the trouble and lead to the appearance of fresh symptoms. Should the psychologist suggest the real source of friction and inefficiency, he is apt to arouse hostility, or to be side-tracked on to dealing with trivial details. Hence his aim, like that of the psychotherapist who is treating a neurotic patient, should be to help the institution to understand its own deeper motives and to achieve a more harmonious integration.

While the present writers are sympathetic with such psychoanalytic explanations of the maladjustments both of individuals and of social groups or institutions, they would stress the subjectivity of these approaches. Just as different analysts diagnose the same patient differently and explain his symptoms (and often ameliorate them) in terms of diverse theories of human motivation, so different "sociatrists" would be liable to analyse a firm's or an educational system's difficulties differently. Obviously, too, very few persons could be trusted to handle such investigations. One shrinks from the prospect of a newly-fledged educational psychologist lecturing his education committee on their unconscious mechanisms which stand in the way of an efficient selection scheme. Nevertheless, some training in abnormal as well as normal sociology should help him, or any other applied psychologist, to realise why scientific selection and guidance schemes are not always immediately acceptable or practicable, to feel his way tactfully and gradually and to bear with his frustrations.

Tests

The principles of testing are discussed in later chapters. Here it is desirable to combat the exaggerated rôle assigned to tests of abilities by many writers, particularly in America. Vocational classification is often described as putting "square pegs" into "square holes," as though it consisted merely in matching mechanically the measured characteristics of the candidates with the appropriate "profile" of job requirements. It is a misleading over-simplification of the aims and methods of vocational psychologists to neglect all the principles in our list except Nos. 4 and 7—Tests and Validation (only admitting Nos. 2 and 5—Personal History Items and Interview Judgments—in so far as they can be treated as additional "tests"). Indeed, nothing is more likely to bring vocational psychology into disrepute, especially if the tests are applied and interpreted by not very highly qualified personnel. It is gratifying to observe that the plans of American psychologists for the rehabilitation and guidance of demobilised recruits assign an important but far from pre-eminent function to tests, and recognise the crucial value of skilled interviewing, of surveying each individual's past record, and of integrating the guidance procedure with training (cf. Scott & Lindley, 1946). In other words they appear to be accepting the view, generally held by British psychologists (cf. Rodger 1939b) that tests should be "servants not masters."

In most jobs such factors as health and physique, interests, attitudes and personality qualities, previous experience and training, and sociological circumstances are more influential than the abilities which can readily be tested. Such factors can, it is true, sometimes be treated quantitatively or measured by appropriate tests (cf. pp. 135–142), but in actual practice they are more often elicited by biographical questionnaire and interview, and weighed up qualitatively.

Long and Lawshe (1947) point out that in 1919 psychologists were expecting tests to be adopted widely in selection for any and every job, but that twenty years later there had been very little advance. Only about 7 per cent. of a large sample of firms in the United States were making any use of tests in 1936 (though there appeared to be a rise in some smaller surveys in 1940). They attribute this partly to undue enthusiasm on the part of testers, partly to undue suspicion among users. In our view the main factors are: (*a*) The extraordinary difficulties of devising practicable tests and of carrying out even preliminary validation, still more of keeping such validation up to date and accommodating the test batteries to alterations in the nature of the work; and (*b*) the importance of the factors other than abilities, just mentioned. Few if any industrial or educational organisations could afford the numbers of highly skilled personnel which would be needed to carry out vocational schemes based wholly on tests. In the 1920's Hull (1928) advocated the construction of some 30–40 tests which between them would measure the abilities needed in 50 of the main occupations. By mechanical methods an individual's scores on such a battery could be appropriately weighted and combined to predict his suitability for all these occupations (cf. p. 104). Only recently has this project been put into effect by the U.S. Department of Labour, Occupational Analysis Division (Dvorak, 1947). Fifteen varied tests, taking about two and a quarter hours, provide measurements of ten main ability factors, and minimum scores on these factors are indicated for twenty groups of common occupations. While the scheme is highly ingenious, we feel that it appears to make vocational classification delusively easy. Moreover, the validity of several of the tests, and of the factors based on them, in predicting occupational success is likely to be low, and no evidence on this point has so far reached us.

The Interview

This topic, too, receives fuller treatment below, and we will merely outline its main functions.

(i) Clarifying and amplifying the data given in the questionnaire and assessing its significance; in particular scrutinising claims to trade experience.

(ii) Providing the interviewee with more detailed information about any jobs in which he is interested or which appear suitable.

(iii) Surveying his interests and attitude, e.g. towards different types of secondary education, to an industrial firm, or to the Service. While it is entirely feasible to devise scales for measuring any one well-defined interest or attitude, or several simultaneously (cf. Vernon, 1938a), yet so great is their variety, and so complex their inter-relations in different individuals, that the more flexible, even if less accurate, interview approach can scarcely be dispensed with.

(iv) Judgments of temperament and character from facial expression, gestures, manner, conversation, and directed discussion of past history and future aims. This is the aspect of interviewing which is most unreliable, and most likely to differ from one interviewer to another. Nevertheless, it can be standardised and improved in many ways, and in the absence of good tests of personality, is almost the only feasible approach to the assessment of this vital factor.

(v) Synthesising and balancing up all the factors such as experience, education, abilities, interests, personality and job possibilities. Here also there is so much scope for subjective judgment that the conclusions reached by poor interviewers may often be less valid than conclusions based, say, on tests alone or on the interviewee's expressed preferences alone. Nevertheless, here also more scientific procedures raise almost insuperable technical difficulties.

(vi) Stimulating the individual's morale. Even if (as may sometimes be the case) the interview is worse than useless, it would still be desirable to include it, at least in Service classification procedures, because it is by far the most "human" of psychological techniques. It shows the recruit that he as a person is receiving consideration, and that the Service is trying to take account of his capabilities and interests. It would be impossible too to convince the Service authorities that either the recruit or the Service itself was getting a fair deal if impersonal methods such as tests and questionnaires alone were used. Similarly, in industry the goodwill towards the firm engendered by sympathetic interviewing is a potent consideration. Finally, tactful persuasion by the interviewer is often needed if the candidate is to accept willingly the job for which he is most suited, rather than the job which he thinks he would like.

Links with Training or Education

Both guidance and selection at the present time tend to be based to too great an extent on a single "cross-section" of the candidates. As pointed out under No. 2, classification procedures among school children can be spread over several years. While the questionnaires used in the Forces attempt to take the past into account, there are obvious advantages in the approach advocated by R.A.F. psychologists, which regards selection and training as parts of one and the same process. The effective placement of a civilian recruit in a Service job involves both his initial allocation and the successful completion of his training (or his re-allocation and re-training). During training his competence for the job is examined, just as it is in selection testing; moreover, each completed stage of training is considered to be predictive of suitability for the next stage, just as is the psychological examination.

Vocational classification in the Services was handicapped by the weakness of the links between the psychological, the training and the education branches. Apart from the fact that they did not always see eye to eye, the effective integration of their aims and techniques was rendered impossible by the shortage of psychologists. The result was that innumerable problems of mutual interest failed to receive the scientific investigation they deserved. Some training courses were largely inappropriate; training devices and visual aids were employed without any proof of their value; old-fashioned examination methods were used which not only probably failed the wrong men, but also provided thoroughly unsatisfactory criteria against which to gauge the validity of selection procedures. Again, more attention might have been paid to the psychologists' view that, as only a limited supply of highly intelligent or well-qualified men was available, certain training courses should be simplified and lengthened for the benefit of the poorer-quality trainees. The naval "transfer" scheme and the Army A.S.C.s, described above, did help as it were to plug some of the leaks, but the most effective schemes of selection-cum-training in the Services were those developed for naval officer candidates, and for R.A.F. air crew.

It is appropriate here to point out the close inter-connection between selection, training, and factors of design and layout in the job itself. When, as so often happens, the work involves needlessly complex perceptual and muscular activities, selection is rendered more difficult and training lengthened. Conversely, the psychologist can often simplify his selection problems by modifying the equipment in such a way that a much larger proportion of the population can manage the job after fairly brief training (cf. Craik, 1945).

The desirability of the conception we are advocating is already generally recognised in education. Vocational and educational guidance should be a continuous process from 11 years to adulthood. At each stage in a child's career the most suitable type of future schooling, or the most suitable vocational plan should be considered. Moreover, the predictive value of the various school examinations should be followed up in just the same way as that of vocational tests. In many industries, again, the selection and training of apprentices, or their transfer to other simpler jobs if they fail to make good progress, are combined in a single department.

Validation

The methods by which educational and vocational classification are carried out are neither new nor mysterious. Psychological tests, interviews and questionnaires are simply refined and standardised versions of the methods to which schools and industries have long been accustomed. Thus, the main difference between psychological and other procedures is that the psychologist insists on validating his instruments and testing his tests, instead of relying on "hunches," or on "experience" and uncontrolled observation. He is aware of the fallibility of human judgment (including his own), of the tendency to jump to conclusions and to generalise from a few instances, to reason in accordance with our sentiments and complexes rather than in accordance with the facts, and to oversimplify the personalities with which we have to deal. Investigations of gradings or assessments of personality traits have shown both how widely different judges of the same person may differ, and how influential is the "halo effect," that is the impregnation of a judge's gradings of specific qualities with his general good or bad impression (cf. Vernon, 1938a). The ordinary man or woman fails to realise, for example, that because a boy cheats in his work at school he is not necessarily destined for a life of crime, and that this peccadillo probably has no bearing at all on his mechanical skill. An untrained interviewer or personnel official is often influenced by even less relevant characteristics—a smiling face, a large jaw, hands in the pockets, a girl's use of lipstick, and so on.

Predictions of vocational suitability are often based not merely on subjective impressions but on such tests as school examinations, or on actual performance in some job. This information too may be extremely misleading unless it is scientifically validated, and proved to measure the qualities needed. A very common and insidious error, of which doctors and psychologists themselves as well as non-scientists are guilty, is the "naming fallacy"—that is the assumption that a test is relevant if it has the same name as a quality involved in the job. The radar operator, for example, needs good "vision," but no one has bothered to investigate whether the aspects of vision tested by the medical officer are actually the same as those used in the work. Applied psychologists in Germany were always very prone to list the hypothetical abilities involved in a job and to assume that, by devising one or two tests of each of these, they could predict job ability. British and American psychologists are generally more empirically-minded, and it is fair to claim that they know more about the predictive value and the limitations of their tests and other methods than do any of the hosts of organisations which conduct educational, professional, trade or other examinations.

Validation and follow-up investigations are carried out with:

(i) Vocational procedures as a whole.

(ii) Each part of a procedure, such as the separate tests, interviews or questionnaires.

(iii) The separate items that compose a test or questionnaire, the aim being to improve the instruments by choosing the best items or revising the ineffective ones.

It is realised too that procedures or parts of them may have satisfactory validity in relation to one type of work, yet not to another superficially similar type. Hence they should be validated afresh for each job, and the process should continually be kept up to date with alterations in the jobs.

Thus, any system of classification which does not provide for the validation and follow-up of its methods, and which is not continually trying to improve these methods in the light of their results, should be regarded as little better than quackery. We entirely agree with Macrae (1932) and others who hold that successful guidance is, in many respects, an art as well as a science, and with Burt's (1942–3) admission that the intuitions of the qualitative or clinical investigator have often led to more original advances in psychology than have the quantitative analyses of the statistician. Nevertheless, it is essential for the vocational psychologist to be trained in scientific techniques of investigation and statistical methods, so that even if he is not himself responsible for follow-up he can realise the importance of statistical checks and can interpret the significance of results obtained by others.

Training of Personnel

Since 1942 the supply of trained psychologists has been quite inadequate to cope with the demands of the universities, the Forces, education authorities and clinics, industrial and other employers, and these demands appear to be increasing. It is essential therefore to hand over much of the day-to-day work of classification to less highly-qualified staff, as was done quite successfully in the Forces. There are, indeed, positive advantages in using officers who know more about the Navy or Army, teachers who know more about education, and personnel officials who know more about industry, than the psychologist himself is likely to know. Moreover, as already indicated, the psychologist prefers to educate organisations to carry out appropriate vocational schemes and to give technical help where needed, rather than to impose a scheme run by technicians. The system of using non-psychological staff worked well in the Navy since the recruiting assistants and P.S.O.s were very carefully selected and trained in the first place, were kept under constant supervision, and were fed with up-to-date information on naval requirements, job analyses and psychological techniques. It was only slightly less satisfactory in the Army because of the very large numbers involved, and because the less effective powers of the qualified psychologists made it difficult to maintain as high standards. In both Services, however, a check was kept on each member of the staff; for the work of a P.S.O. who interviews and makes employment recommendations is a technique which requires validating just as much as is an intelligence test. It was often possible to eliminate the weaker ones or to transfer them to duties for which they were more suited. Problems both of overwork and of boredom were often serious, but could be avoided under conditions which reduced the testing and interviewing load and allowed the personnel to be trained for more responsible and more varied tasks, such as follow-up, documentation, job analyses, and research investigations.

Large-scale applications of vocational classification in schools, such as the Birmingham and Ayrshire experiments, have relied mainly on the part-time work of partially-trained teachers, and short courses given by the N.I.I.P. have helped to provide moderately well qualified careers masters and mistresses for schools all over the country. But there are very grave dangers in such schemes being undertaken by insufficiently skilled personnel, dangers not only to the pupils or employees who are misdirected, but also for the reputation and progress of the whole of vocational psychology. At present an employer or education authority has no easy means of distinguishing a competent person from a quack. A degree in psychology or possession of Associateship of the British

Psychological Society does not necessarily show ability to undertake selection or guidance. Nor, on the other hand, are the many persons who acquired valuable experience in the Forces and carried out excellent classification work necessarily competent to organise civilian schemes on their own. Presumably what is needed is a central training institution conferring diplomas which would be as widely recognised as the London School of Economics' Mental Health Certificate for psychiatric social workers (cf. Rodger, 1944). It would be desirable too for such diplomas only to cover a limited period of years, a refresher course and a re-examination in knowledge of techniques and jobs, and in interviewing skills, being required before they were re-issued. Fortunately, psychologists are generally aware of the present unsatisfactory state of affairs, hence there is reason to hope that steps will be taken both to increase the supply of trained workers and to assess and maintain their standards of efficiency.

A further problem which has not yet been studied at all fully, is the ethical one of the use to which information amassed by psychologists or their assistants may be put. Often such information deserves to be treated with as great confidence as that received by a doctor or solicitor. It would be intolerable, for example, if children's test scores, or material elicited in clinical interviews, were passed on to employers. We did not, in the Forces, even allow the information we possessed about recruits to be handed to other Government departments without the recruits' permission, since it had been collected not voluntarily but under conditions of Service discipline. At the same time the psychologist is not in the position of a doctor who can state: "X is not fit to do this type of work, but I will not disclose what is the matter with him." The solution to such problems appears to lie, not so much in the development of still more reliable and valid techniques, as in the establishment of vocational psychology as a profession such that the judgment of a psychologist with appropriate qualifications will be accepted as readily as those of his medical or legal colleagues.

CHAPTER VII

THE VALUE OF VOCATIONAL CLASSIFICATION PROCEDURES

Abstract.—Validation of tests or other procedures necessarily involves such statistical concepts as correlation coefficients, partial and multiple correlation, and reliability. Procedures may be compared with a variety of criteria of proficiency, i.e. of success or failure at the work, including:

(*a*). Objective records of output, accidents, etc.
(*b*). Differences between groups of known proficiency.
(*c*). Results of examinations, trade tests, etc.
(*d*). Merit ratings or other subjective gradings.

Examples of these types are listed. Some of the difficulties in collecting reliable information on proficiency under Service, or under industrial, conditions are pointed out. Large numbers of cases are needed for adequate validation studies, and the variability of results in different studies should be checked. Proficiency often depends on a great many factors besides the suitability of the individual workers, and these must be controlled as far as possible. The workers whose success or failure is followed up often constitute a selected group, and this seriously distorts the results obtained. Thus, in judging the value of vocational classification these technical difficulties must be borne in mind.

Pre-war research has demonstrated the effectiveness of vocational guidance as developed by the N.I.I.P.; but substantiation of student counselling and child guidance procedures is so far less adequate. Statistics of the proportions of satisfactory trainees or of officers in the Services, who have been selected by psychological methods, do not suffice to prove the value of these methods. But several studies are cited where recruits so selected were found to be more successful than those selected by older methods. Though the follow-up, and assessment of the ultimate proficiency of, officers are particularly tricky, some fairly satisfactory results were obtained on the validity of W.O.S.B. procedures as a whole. Individual board members showed considerable variations in skill,

and much more investigation is needed (and is now being carried out) into the reliability and validity of the separate parts of board procedures.

The importance of validation was pointed out in Chapter VI. Inevitably the subject is a somewhat complex and technical one, but we will attempt to describe its main principles and difficulties as simply as possible. Thereafter we shall present some of the evidence for the value of vocational adjustment procedures as a whole. Certain essential concepts such as correlation and reliability have been discussed elsewhere (cf. Vernon, 1940), and are fully treated in most statistical textbooks. Thus brief definitions will suffice here.

Correlation.—When any two sets of scores or gradings are compared, for example results on a selection test and subsequent degree of proficiency, the closeness of agreement can be expressed as a correlation coefficient ranging from + 1·0, meaning perfect agreement, to 0·0 meaning no agreement at all. If those who do well at the test tend to do badly at the job, the coefficient will be negative, though this rarely occurs. More often coefficients are low positive (+ ·2 to + ·3), moderate (+ ·4 to + ·5), or fairly high (+ ·6 to + ·7), though far from perfect. A note on the implication, or concrete significance, of the size of correlation coefficients is appended at the end of this Chapter.

Partial Correlation may be used when it is desired to find whether some additional test, examination mark or grading contributes anything to the prediction of some criterion, say job proficiency, which is not already covered. Suppose, for example, pupils are selected for secondary education solely by an English examination, it is likely that the addition of an arithmetic examination would cover rather more of the ground, provided that English and arithmetic marks do not themselves overlap very closely. If the pupils' marks on the two subjects correlate perfectly, then obviously either will predict as efficiently as the two combined. The partial correlation of arithmetic shows its extra contribution when English is, as we say, "held constant."

Selection is generally based on several tests, etc., for example, English, arithmetic and an intelligence test, and their combined predictive value is shown by their *multiple correlation* coefficient. As some tests, considered singly, are likely to be more valid than others, the better ones should, of course, receive more weight. By means of statistical analysis a multiple regression equation can be found showing the most appropriate weighting for each test, and by summing these weighted scores the highest possible multiple correlation is obtained.

Reliability refers to the stability of test scores, or their trustworthiness, quite apart from whether the test measures any particular ability or predicts any job proficiency. If testees' scores alter considerably on taking the test a second time, or on taking a second exactly parallel test, the reliability (shown by the correlation between their two sets of scores) will be low. Such an untrustworthy test is unlikely to possess good validity either. When both a test and a criterion have reliability coefficients of ·6, the maximum possible validity of the test will be ·6 instead of 1·0. A good test of adequate length should have a reliability exceeding ·9, and certainly not less than ·75*. More often the criteria against which selection procedures are validated show poor reliability, as when the correlation between judgments of the efficiency of the same men by two instructors reaches only + ·5.

Statistical Significance.—There are such big differences between different individuals that results based on only a few cases—whether average test scores or correlations, etc.—are apt to vary also, or to be untrustworthy. Thus, in one group of, say, six men there might be perfect agreement between an aptitude test and proficiency, in another group of the same size no agreement at all. But the extent of such chance variations can be calculated and the trustworthiness of a result determined. A correlation coefficient is said to be statistically significant when we have proved that it would be most unlikely to have arisen by chance. Similarly the difference between the scores of two groups, or between the correlations obtained in two groups, can be shown to be significant and reliable, or not so.

Types of Criteria in Validation

A great variety of criteria have been used in validating guidance and selection procedures. The following classification is based on articles by Farmer (1931) and Stott (1939, 1943), Viteles's book (1932), and on experience gained in the Forces.

---

* This account is greatly over-simplified, since reliability coefficients depend on numerous factors, such as the population tested, and the method of calculation, as well as on the test itself. Moreover, a test with quite low reliability may be quite worth using if it makes a significant contribution to the validity of a battery. But this is not the place for a full discussion (cf. Guilford, 1946; Cronbach, 1947).

A. *Objective Records of Individual Performance*

Selection tests have been compared with the quantity of output in industry, or with objective measures of the quality of the work, with words per minute transmitted or received by telegraphists or taken down or typed by stenographers, with hits on a target in rifle-shooting, and so forth. Indirect measures of proficiency include wages on a piece-rate job, accidents among operatives or motor drivers, breakages or spoiled work, and savings in the cost or time of training employees. The numbers of changes of posts, or the length of tenure of posts, have been used in assessing the value of vocational guidance.

While such criteria are generally the most satisfactory because they are the most clear-cut and the least affected by subjective judgments of employers or instructors, yet they often show rather poor reliability (hence records may need to be collected over a considerable period), and are often distorted by factors not under the control of the worker, e.g. breakdowns in machinery or Trade Union restrictions. It is seldom, too, that comparable measurements can be obtained on a sufficiently large number of cases.

B. *Differences between Groups with Known Characteristics*

Men engaged on a given job, and presumably for the most part competent at it, may be contrasted with others not so engaged. Thus, a mechanical aptitude test should differentiate between mechanics and clerks. Advancement is a useful criterion, as when officers, instructors or promoted men are contrasted with beginners or "other ranks." "Drift to recommended job" is explained below. An example of an indirect measure is operating costs such as the saving in current on a tramway system after the institution of selection.

C. *Results of Theoretical or Practical Examinations*

School or technical examinations, trade tests, tests of elementary training among infantry, constitute examples. In lengthy training courses with several stages, the stage that a trainee reaches without failing may be used. Various combinations of examination marks may be derived, e.g. by factor analysis (cf. p. 168).

In all of these the judgment of one or more examiners, instructors, etc., enters either in the marking or grading of the candidates' work or—as in "new-type" tests—in the setting of the questions (cf.

Vernon, 1940). Hence variations of standards and the prejudices of different examiners inevitably affect the results.

D. *Gradings or Assessments*

This includes gradings based on observations of the man rather than of his performance at particular tasks. Merit ratings in industry or specially collected assessments of the efficiency of officers or men in the Forces are the main examples. Such judgments are extremely liable to be biased by the social qualities, or conformity to discipline, etc., of the people being graded. Many forms of rating scale and questionnaire have been developed in an attempt to reduce the effects of halo, variations in the judges' standards, and other defects (cf. Vernon, 1938a). Some of the methods adopted in the Forces are described below.

Other criteria under this heading include the award of military decorations, the development of psychiatric breakdown, and self-ratings of satisfaction with their jobs by people given vocational guidance.

Naturally there is much overlapping. For example, trade tests (C) are often marked objectively and may, therefore, fall under (A). In many examinations, too, the examiner knows the examinees and is influenced in his marking of their work (C) by his general impression (D). Another classification which cuts across this is "training *v.* operational." All types of criteria may be applied during or at the end of training, or else to fully trained men working at the job.

The Collection of Validatory Data

Some books on applied psychology give the impression that validation is a simple and straightforward matter of correlating selection test results with a criterion of proficiency. Actually, it is very rare either in industry or education that a satisfactory, "ready-to-wear," criterion is available. Under conditions of practical life, as contrasted with the controlled conditions which the psychologist imposes when working in his laboratory, there are innumerable disturbing influences which must be taken into consideration and, as far as possible, corrected.

Straker (1944) has given an illuminating description of the collection of data in the Navy. Least information is usually available about the most interesting cases—those who fail in their training or at the job. The psychologist should if possible interview such men and their instructors in order to pin down the causes of failure, whether medical, disciplinary, or lack of ability or interest. Some training schools get rid of a number of trainees on arbitrary and inadequate grounds before the training course even starts, hence he should work on lists of men put forward for training rather than on the school's own class lists. Next he may find that several men have been "back-classed," i.e. transferred to a later class in order to increase the effective length of their training, and again the reasons may be varied—medical or compassionate as well as inefficiency. Training courses often have several sections which are separately marked, and the psychologist must decide which of these are essential for his purpose. The mere final pass or fail results are unlikely to be adequate for statistical analysis.* When a particular stage of training is conducted in more than one establishment, or by several different instructors, it is important to study the comparability of the sets of results. Again the psychologist's plans are liable to be upset by sudden alterations in the length of the training course, or in the syllabus, e.g. with the invention of some new piece of equipment, or by the transfer of the training to a new school. Although the numbers under training at any one time may be small, he cannot usually wait to collect proficiency data on large groups. He must select and advise recruits as best he can, on the basis of past experience of similar jobs and on his own judgment, and improve his techniques as he goes along. All these difficulties have their parallels among school children (cf. McClelland, 1942), and among apprentices or trainees in industry.

Success at training is obviously a less satisfactory criterion than success at the job, but nevertheless usually has to be employed. Training may last only a few weeks but often extends to one or more years among skilled tradesmen and professional workers, e.g. doctors. In such cases the psychologist should try to follow the trainees throughout, but is forced to guide his methods chiefly by results obtained in the early stages. Moreover, he has to select men who will pass training courses even if he is convinced that job requirements differ considerably from training requirements. Much evidence was collected, especially in the Army, as to the inappropriateness of some training courses. For example, many recruits from civilian engineering trades could not be recommended for corresponding Army trades since they were too poorly educated to learn the required theory. Yet, in the field, these same men were transferred into tradesmen, without any training, and managed the work satisfactorily.

---

* The statistician can use any kind of results, such as rankings from highest to lowest, or Excellent, Good, Average, Fair and Poor grades, or merely Pass *v.* Fail, provided the numbers are large enough. But an extended distribution of scores is the most convenient, and most economical of numbers (cf. Vernon, 1946b).

When gradings (Type D) have to be used rather than marks (Type C), it was found best to formulate highly concrete questions which forced the grader to assess the performance of each man in his training or at his job, instead of giving an all-round opinion or judgments of general personal qualities. Useful questionnaires for following up officers were developed by W.O.S.B. psychologists, containing some twenty to thirty items each with two to four possible answers*. A similar but briefer questionnaire for other rank investigations is reproduced here. Though the filling up of such a form for a number of men appears formidable, graders find it easy to use, since it is couched in their own language, and feel more confident than they do in grading general traits. It is best done under supervision but can be sent out and returned by post when the men to be assessed are scattered in many different units.

FOLLOW-UP QUESTIONNAIRE

PART I.—GENERAL SOLDIERLY QUALITIES

A. 1. However great the stress of battle, he never showed signs of cracking.
   2. He stood up quite well when things were unhealthy.
   3. He needed nursing when the going was tough.
   4. He was not much use in battle.

B. 1. In action you could always depend on him to do the right thing without being told.
   2. When not under immediate command, he usually kept his head.
   3. He always needed to be led and told what to do.

C. 1. Hard conditions tended to get him down.
   2. He accepted bad conditions cheerfully enough.
   3. He helped to keep up the men's spirits when conditions were bad.

D. 1. He is rather a disturbing influence.
   2. He gets on all right with the others.
   3. He is always very popular in any group he is with.

E. 1. He always kept alert and vigilant, even when direct contact with the enemy was not expected.
   2. He was reasonably vigilant.
   3. He was inclined to become slack after a period of strain.

F. 1. He is an intelligent soldier who can adapt himself to novel or unexpected situations.
   2. He is bright enough, and seldom at a loss.
   3. This is rather a dull man who gets stuck when he meets some unforeseen snag.

---

* W.O.S.B. follow-up was influenced, like W.O.S.B. selection, by Field Theory. Hence it avoided asking for judgments of independent traits. Actually, however, the same conclusions as to the advantages of concrete, graphic rating scales had been reached by Behavioristically-minded American psychologists in the 1920's (cf. Vernon, 1938a).

G. 1. His attitude to authority and to correction is apt to be "difficult."
2. His attitude towards his seniors is that of an ordinary good soldier.
3. He is exceptionally loyal and co-operative towards his officers and N.C.O.s.

OVERALL ASSESSMENT

Outstanding	Very High	Satisfactory	Passable	Not up to Standard
(Best 3)	(6)	(12)	(6)	(3 Least Satisfactory)

PART II.—PROFICIENCY AT HIS JOB

Main Job........................ Alternative Job................

Operational Experience : Extensive? Moderate? Slight?

H. 1. He is outstandingly effective in handling tools of his trade (arms, tools, equipment, etc.).
2. He is competent and effective enough.
3. He is rather ineffective.

J. 1. His heart has never really been in his job.
2. He is interested in his job.
3. He is a real enthusiast about his job.

K. 1. Physically he is well up to the job.
2. He is handicapped at his job by physical shortcomings.

L. 1. He has a flair for improvising (with tools, materials, etc.) in an unexpected difficulty.
2. He is reasonably good at making the best use of what is to hand when things go wrong.
3. He is lost without the usual tools, materials, etc.

M. 1. He has no eye for ground and cover and is apt to make clumsy mistakes.
2. He makes reasonably intelligent use of country.
3. He has a real "poacher's instinct" for using ground and cover.

OVERALL ASSESSMENT

Outstanding	Very High	Satisfactory	Passable	Not up to Standard
(Best 3)	(6)	(12)	(6)	(3 Least Satisfactory)

A distinct method developed by Admiralty psychologists in 1942 might be termed the clinical method, since the graders are not required to give any verbal, numerical, or other gradings. A conference is held with the officers or others who have had most to do with the men, and a general discussion is held to bring out the strong and weak points of each man as a whole. The discussion is kept on as concrete a level as possible; such questions are asked as: "Which men would you choose to take with you on a small landing party, and why?" The psychologist thus tries to disentangle the efficiency or other relevant qualities of the men from the less relevant social-emotional qualities which so often bias the graders' judgments. For example, cross-examination may reveal that a recruit would be considered excellent at his job, but that he is an ardent communist. From the discussion the psychologist decides which men are, say, outstanding, above average, fair or poor in respect of the qualities that concern him. He can also judge whether some of the raters are too prejudiced or too little acquainted with the men for their cases to be worth including in the investigation. He should, of course, be ignorant of the men's test scores or other selection data, so that his gradings will be entirely impartial. Both the questionnaire and the clinical approaches demand a thorough knowledge of the job, so that the questions may be framed around its essential features.

Often conference and rating scale methods are combined. The raters of several groups of men can be present simultaneously, since this helps to familiarise them with the procedure. But if there are two or more raters for any group of men (a highly desirable condition), the most junior in rank should be asked for his opinion first, or they should be interviewed independently.

Clearly the collection of good follow-up evidence requires an experienced and tactful investigator. For example, some officers might resent the imputation that any of their men are poor or inefficient, but may be persuaded to discriminate if asked, "Which men need to improve?" W.O.S.B. psychologists point out an important difference between "out-group" and "in-group" judgments. Trainees who will shortly pass on to other units are regarded by their instructors or temporary officers in a very different way from men belonging to an officer's own company or crew.

Treatment of Validatory Criteria

Only a few of the complex technical problems raised by validation can be mentioned here. Often the apparently poor showing of vocational procedures is largely due to unreliability or inappropriateness of the criteria. Accident rates, for example, are very unreliable. Even if collected over as long a period as two years, the correlation between the rates in the first and second year is generally no higher than + ·4. The untrustworthiness of ordinary written examinations is well known. Two parallel one- or two-hour papers in the same subject usually correlate ·6 or ·7, and the agreement between two examiners marking the same papers may be no higher. Practical or oral tests are generally poorer still since they can only sample a small fraction of the candidates' skill or knowledge. Merit ratings on personal traits or general efficiency are commonly found to yield reliability coefficients around ·5. However, by combining several sets of observations, marks or ratings, a sufficiently precise or stable criterion can usually be achieved*.

---

* When the candidates take half a dozen or more theoretical or practical examinations, or are assessed on several characteristics, factor analysis is particularly useful for determining the underlying qualities, and from the results a more clear-cut criterion can be built up. Such analysis will show, for example, whether it is legitimate—as most schools and training institutions do—to add all the

Even when criterion scores or grades are statistically reliable, they may or may not be representative of the proficiency in which the investigator is interested. Rifle-shooting scores, for instance, obviously do not coincide with all-round proficiency of infantrymen. In no practical job are marks on purely written examinations adequate, and they need to be supplemented with assessments if practical tests are not available. Thus the investigator generally requires to make at least a rough job analysis, and should consult instructors or other experts, in order to interpret the significance of the proffered criteria.

Human beings vary so widely, and their success or failure at a job depends on so many factors, that no sound conclusions as to the usefulness of a vocational procedure can be derived from a few cases. One of the main reasons for the slow progress of vocational psychology has been the small number of persons in any one civilian job available for investigation. In education and in the Forces psychologists have been more fortunate. No fixed rules can be laid down, but a study of fifty cases or less will usually suffice only to give some rough indications to the psychologist himself. A minimum of two hundred is desirable for proving the value of some selection technique, and eight hundred for providing detailed conclusions, e.g. on the scoring of items in a questionnaire or the appropriate weighting of a battery of tests. In order to render a conclusion twice as secure, or to halve the influence of chance variations, it is always necessary to quadruple the cases. But much depends on the care with which the data are collected, and on the degree of refinement or coarseness of the measures employed.

As far back as 1924, Thurstone pointed out that random fluctuations due to the "luck of the draw" are much less serious than variations in the results of different groups brought about by conditions which the investigator cannot readily control. The agreement between some procedure and proficiency may differ in different firms, schools, or Service units, because the groups are differently motivated, e.g. one tester obtains better co-operation than another, the training or education differs, the attitudes and general morale of the groups vary, the jobs at which proficiency is assessed—though nominally the same—differ sufficiently to demand different patterns of abilities, etc. Time and again an investigator has

marks together as measuring a general proficiency factor, or whether it is better to regroup them into two or more sets and to validate the selection procedures against each set separately.

obtained very promising results in a preliminary validation experiment conducted under carefully controlled conditions, perhaps on a small number of cases, only to find the value of the procedure very much reduced when he follows up a larger and more miscellaneous sample. Indeed, a widely accepted rule among American psychologists is to insist on double validation. If the best combination of a certain battery of tests is worked out from the results of one group, this combination should be tried out and followed up in another group before it is put into routine practice. Such statements as, "Had the following pass-marks been applied on the following tests, X per cent. of the men selected would have proved successful," should be abjured.

It is becoming recognised in educational psychology that no experiment carried out, even on large numbers, in a single school is adequate, and that variations in results between several schools, rather than variations between individual pupils, should be explored by means of analysis of variance techniques (cf. Lindquist, 1940). We have hardly begun to apply this conception in industry and the Forces. It would, however, undoubtedly be far more useful to carry out validational investigations on four groups of fifty each, in four firms, training schools or units, and to analyse the extent of the differences in these four contexts, than to make one investigation of two hundred cases. The latter will only tell us that our procedure is valid if applied to other exactly similar samples under identical conditions, whereas the former will tell us whether our results are sufficiently consistent to be applicable to other groups where conditions may vary as much as they did among the four actually studied.

Allowance for Group Differences

Unfortunately the statistical techniques for the type of investigation just mentioned are elaborate and not widely known among psychologists. And as it is very rarely possible to collect hundreds of cases selected, trained, and graded for proficiency under identical conditions, many tricky problems have arisen in combining the results of numerous small groups. Take the common instance where the criterion consists of examination or trade test marks awarded at the end of training, or to pupils at the end of a school year, where the trainees or pupils have received their instruction in several more or less parallel classes. Such marks depend on:

(1) The examinees' ability, knowledge and skills.
(2) Their interest, industry and other personality qualities, including "examination nerves."
(3) Relevant experience or knowledge prior to training.
(4) Length of training course.
(5) Goodness of instruction or teaching.
(6) The examiners' standards of marking and personal idiosyncrasies.
(7) Miscellaneous environmental, health and other conditions.
(8) Chance factors—the luck of the questions set.

While this analysis applies primarily to criteria of Type C, similar sources of variation occur with other types. Thus, in Type D it is seldom possible to get large groups assessed by the same judge, hence the various judges' standards of grading and their varying interpretations of the qualities to be graded tend to produce large group differences. Under Type A, weather conditions obviously affect the rifle-shooting scores of groups of men who fire on different days; accident rates depend on the exposure each group has undergone, and so on. Now selection procedure only attempts to predict Nos. 1 and 2 on our list, and to take No. 3 into account. It is essential therefore to keep Nos. 4 to 8 as constant as possible in all the groups followed up, if the true agreement between the predictions and the results is to be obtained, or else, as already indicated, the investigation must be extended to several groups and the consistency of the findings analysed. In collecting follow-up data as full information as possible must be obtained under headings 3 to 8. When these factors operate randomly, neither helping nor hindering groups of initially good or poor quality, and when they produce variations of not more than about 10 per cent. in the marks, then it can be proved that their effect is merely to reduce the validity coefficients slightly (cf. McClelland, 1942). But more often they operate selectively, for example, the initially poorer groups get poorer training, or alternatively they are provided with better instructors. This greatly distorts the correlations, occasionally raising them spuriously, more often reducing them considerably. Similarly when an officer or instructor assesses the proficiency of his men more highly than does a second judge rating a second group, it may be difficult to tell how far the first group is really superior in quality, how far the difference is due to No. 6 on our list, and there is usually no means of allowing for its effect on the correlations*.

Selectivity

Psychology resembles the other sciences in that the only sound way it has of determining the effect of a certain condition is to observe the results of applying that condition, all other factors being kept constant. Thus the ideal method of validating a vocational procedure is to compare the results in a group subjected to this procedure with the results obtained in another group precisely similar in all respects except for the absence of the procedure. This often raises insuperable practical difficulties. One may compare output or examination results before and after the introduction of a procedure, but almost always the training given to the second group, or the job itself, will also have altered, or other changes in the examiners, the morale of the workers, etc., will have occurred. If gradings are used, they will depend more on the graders' opinions of the procedure than on the efficiency of the men. An employer's satisfaction or dissatisfaction with the men sent him at different periods does not constitute scientific evidence. Every effort should be made then to follow up simultaneously a group subjected to the procedure and a control group; and when subjective judgment enters into the criterion, the examiners or graders should not know which men belong to which group.

Naturally, however, neither the Services, nor employers or education authorities, can usually afford to accept such unselected trainees, when it is certain that a large proportion of them will fail, and the psychologist has to make do with a very unsatisfactory alternative. In effect he grades his candidates, at the time of adjustment, into more and less suitable groups and subsequently finds whether the former are superior to the latter on the criterion. Thus he is forced to deal with selected, and unduly homogeneous, groups. He can study the value of, say, tests for secondary school pupils or mechanics only by following up the careers of pupils or mechanics who were originally chosen largely on the basis of these tests. Not only are the differences between his more and less suitable groups likely to be smaller than the differences between all his selectees and a control group—hence his validity correlations are greatly reduced—but also all other correlations tend to be distorted to a varying extent by this selectivity. Suppose, for example, he selects men on the basis of previous education, but applies selection tests with a view to choosing the best tests and substituting them later for the education: he is almost sure to find the tests agreeing better with proficiency than does education, and yet their agreement will certainly be poorer than it would have been had the men been entirely *un*selected. In other words, any test or selection procedure *not* used for selection appears to give better predictions than procedures which *have* been used, yet no procedure can display its full predictive value unless—a rather unlikely contingency—it is quite independent of, or uncorrelated with, the procedures actually employed. Again it follows that the better the selection scheme already in operation, the poorer will its validity appear to be when a selected group is followed up. In one experiment in a new mechanics branch in the Navy, the first 300 men sent for training were almost unselected since little was known about the job requirements. With the next 300 more information was available and standards were raised. With the next 500 there was still further improvement in the selection procedure. But the correlation coefficients between T2 and course results in these three groups were ·69, ·53 and ·38 respectively, and all the standard naval tests showed similar declines in validity (cf. Vernon, 1946b).

---

* McClelland does, indeed, provide a method of scaling the results from numerous groups which brings them to a common standard, but this is applicable only under conditions which never operated in the Forces, namely the existence of some test taken by all the groups which correlates very highly with their ability.

There are, it is true, statistical formulae which can theoretically correct for this selectivity and give the correlations that should have been obtained in an unselected group (Burt, 1943b). Their weakness is, however, that the precise basis of selection is seldom known. Some psychologists or P.S.O.s may have relied largely on test scores, others on previous education or experience, others on judgments of temperament and interests. The formulae only work when all the candidates have been chosen on a particular test or battery of tests, so that the exact amount of homogeneity (restriction in range of ability) can be determined*. It often happens, for example, that P.S.O.s allocate to a certain job men with low test scores, but with compensating qualities of keenness or relevant experience, and when these men do well at the job, they reduce the size of the correlations between the tests and the criterion. No method of correction is adequate in this situation.

---

* They assume too that all the variables involved give normal distributions of scores; this is not necessarily true of all selection tests and is usually untrue of such factors as experience.

In conclusion, it may be seen that validation only approximates to being an exact science. In most practical situations the validatory results are affected by so many uncontrolled influences that a large measure of statistical judgment is needed in arriving at sound conclusions. Often when these results seem disappointing this may not be the fault of the psychologist or of his methods. Sometimes also when they seem rather favourable they need to be accepted with caution; the samples may have been too small, or unduly heterogeneous, or other defects may be present. We must hope that, as the principles outlined above come to be more widely understood, it will be possible to plan better experiments which will yield more clear-cut answers to validation problems.

Previous Validatory Studies

It is difficult to find any good instances in the literature of scientific validation of vocational *selection* procedures which employ interview and other techniques besides tests, though there are numerous claims, not very well supported, as to the benefits of introducing psychological methods into industrial organisations, and thousands of studies of tests. Vocational *guidance*, on the other hand, having been conducted largely by academically trained psychologists, has always laid more stress on thorough validation. A useful summary of British investigations to date is given by Stott (1943).

Of the cases who had received full-scale individual guidance from the N.I.I.P., 463 were followed up three to five years later. From replies to a questionnaire or interview, the numbers who were "successful and happy in their work" were judged. Of the 345 cases in jobs which accorded with the Institute's recommendations 92 per cent. were successful, whereas of the 118 in "non-accordance" jobs only 57 per cent. were successful. Considering that individuals seeking guidance from the N.I.I.P. tend to be "difficult" cases, whose vocational suitability is very uncertain, the 92 per cent. success is highly gratifying. So many conditions may alter in three to five years that a larger percentage could scarcely be expected.

Investigations which conform more closely to what we have called classification, rather than guidance or selection, are those of Rodger (1937) in a Borstal institution and of Hunt and Smith (1945) at Birmingham. Rodger compared assessments by instructors of the efficiency and progress of 200 youths allocated to work parties in the usual manner by their housemaster, and 200 allocated by himself on the basis of tests and short interviews. Of the former 45·6 per cent. and of the latter 69·5 per cent. were given Grade A reports. While the superiority of the psychologist's cases is not very striking, it is probably reduced by the unreliability of the instructors' gradings, and other factors.

In the largest Birmingham enquiry, 1,639 children were followed up over two years and 603 over four years, roughly half being boys, half girls. Half of these had been given guidance by psychologically trained teachers on leaving school, the rest—a control group—had only received advice at ordinary employment conferences. Numerous criteria were employed, including employers' gradings and length of tenure of posts, all of which gave favourable results. The most interesting figures are listed in Table I. Among the "guided" cases, 90 to 93 per cent. of those in recommended jobs state that they are satisfied, and only 26 to 33 per cent. of those in "non-accordance" jobs. Of the controls, however, the cases who followed the employment conference's advice are if anything less satisfied than those who did not. Many more guided cases retain accordance than non-accordance jobs over a long period, but the controls show practically no difference. Finally, it is noteworthy that as time passes a larger proportion of guided cases are found in accordance posts; that is they show a "drift to recommended job," which is absent in the less skilfully advised controls.

TABLE I.—RESULTS OF BIRMINGHAM VOCATIONAL GUIDANCE EXPERIMENT

	Psychologically Guided Group		*Control Group*	
	Accordance jobs	*Non-accordance jobs*	*Accordance jobs*	*Non-accordance jobs*
	per cent.	*per cent.*	*per cent.*	*per cent.*
Satisfied with job at end of 2 years (self-ratings) .	90	26	64	76
Ditto at end of 4 years .	93	33	64	78
Retained first job over 2 years	56	11	37	33
Ditto, over 4 years . .	46	11	27	26
In accordance jobs at end of 2 years	87	—	52	—
Ditto, at end of 4 years .	92	—	47	—

Student counselling procedures in American schools and colleges do not yet seem to be so well validated. Thus, Williamson (1936) summarises several investigations in which arts and engineering students who were counselled by faculty members obtained no better college grades than matched controls. But it is improbable that these counsellors had received any psychological training. In a later research (Williamson and Bordin, 1940), 80 per cent. of students who had received counselling and 66 per cent. of a control group were assessed, one year later, as emotionally well-adjusted. The difference is rather small, yet statistically significant. There have been numerous follow-up studies of child guidance, though few have taken the essential precaution of including a control group in order to discover what proportion of maladjusted children improve *without* receiving guidance. Nevertheless, C. R. Rogers' book (1939) shows that the effectiveness of different types of treatment, e.g. boarding out in foster homes, is now receiving scientific study. Burt and Lewis (1946) briefly mention the application of analysis of variance to the evaluation of different modes of treatment of delinquents and neurotics, one result of which seemed to be that the personality of the therapist is more influential than the type of treatment he adopts or the psychological theory on which it is based. Typical of many interesting investigations of the effects of different systems of upbringing on the personalities of the children is that of Baldwin, Kalhorn and Breese (1945), proving that a democratic rather than autocratic atmosphere in the home produces greater intelligence, originality, curiosity and tenacity, and that over-indulgent or rejectant attitudes among the parents have unfortunate effects.

The Value of Personnel Selection Procedures as a Whole in the Forces

Information on the failure rates of practically all recruits trained for different A.T.S. jobs in 1944 was collected, and it was concluded that 94 per cent. had been successfully allocated by P.S.O.s to jobs they could manage. A similar enquiry covering over 8,000 Army recruits, in several representative Arms, yielded failure rates averaging less than 5 per cent. Such figures, however, do not constitute evidence, since the equivalent failure rates for recruits posted before D.S.P. was established are not known. Also the quality of recruits may have risen or the standards of the training schools may have been lowered. If they show anything at all it is that the Army and A.T.S. authorities were mostly converted, by 1944, to approval of personnel selection.

Much more convincing was a comparison of several groups of auxiliaries selected by P.S.O.s with other groups selected by other methods and trained simultaneously. Their respective failure rates, given in Table II, show great improvement attributable to personnel selection. Special operators were particularly striking, since improvements in selection procedure were introduced in several stages, and with each alteration there was a large drop in the failure rate.

TABLE II.—FAILURE RATES OF A.T.S. TRAINEES SELECTED BY DIFFERENT METHODS

Category	*No. Selected by Old Methods*	*Per cent. Failed*	*No. Selected by P.S.O.s*	*Per cent. Failed*
Drivers	124	30	1,004	14
Clerks	128	11	592	4
Special Operators	420	60	130	7
Operators, Wireless and Line	217	7	187	0·5

War Office and Admiralty psychologists were especially interested in the selection of tradesmen and mechanics, and collected some of their most striking evidence in this field. Hotoph compared the failure rates on training courses of some 10,000 Army tradesmen selected by four different procedures during four months of 1942, with the results listed in Table III.

Apparently the P.S.O.s' selectees had only about half the overall failure rate of men selected by ordinary Army methods and a third that of men selected by the Ministry of Labour. However, the comparison is somewhat unfair in that more of the third than of the

TABLE III.—FAILURE RATES OF ARMY TRADESMEN SELECTED BY DIFFERENT METHODS

Method	*No.*	*Overall Failure Rate*	*"Adjusted" Failure Rate*
		per cent.	*per cent.*
Nominated by C.O. or technical officers	3,566	17·0	19·2
Nominated at own request . .	3,176	16·6	19·6
Called up by Ministry of Labour as semi-qualified tradesmen . .	911	27·1	19·4
Selected by P.S.O.s . . .	2,201	8·7	11·1

fourth group happened to go to trades where failure rates were abnormally high. An attempt was made to adjust for this and to equalise the numbers of each group in each trade, with the results given in the last column. It will be seen that the P.S.O.s' selectees are still the best, though less strikingly so, and that Army and Ministry of Labour cases achieve much the same rates. It should be remembered that Army P.S.O.s had little psychological or industrial training and did not, at that time, use any standardised trade tests or tests of trade knowledge. Hence the main reasons for their success must lie in the careful checking, at interview, of the candidates' previous experience, and in the rejection of candidates who failed to get fairly high scores on intelligence and other group tests.

In the Fleet Air Arm over 16,000 mechanics and fitters were followed up. Over 6,000 were selected by standard methods before psychologists played any part, and their overall failure rate on course was 14·7 per cent. When selection was undertaken by P.S.O.s, working under a psychologist, not only did the failure rate for 10,000 men drop to 4·7 per cent., but also they extracted a much larger proportion of trainees from the available naval recruits without denuding other mechanical branches which were likewise making large demands at that time. Table IV shows that improvement was obtained in every category. While the comparability of the training courses and examinations in the two periods cannot be ensured, careful enquiry revealed no grounds for supposing that

TABLE IV.—FAILURE RATES OF FLEET AIR ARM MECHANICS AND FITTERS SELECTED BY DIFFERENT METHODS

Category		*Old Method*		*Psychological Method*	
		No. of men put on course	*Per cent. Failed*	*No. of men put on course*	*Per cent. Failed*
Air Mechanics	Airframes	1,333	16·8	2,234	6·9
	Engines	1,219	21·2	2,218	5·6
	Ordnance	902	7·3	1,596	0·6
	Electrical	832	16·1	1,969	4·9
Air Fitters	Airframes	953	10·1	723	4·3
	Engines	890	15·6	743	5·4
	Ordnance	130	13·1	199	2·0
	Electrical	271	9·2	326	1·8
Total		6,530	14·7	10,008	4·7

there was any relaxation of standards during the period when psychologists took over. It has been calculated that this reduction in training wastage resulted in a financial saving of fully £100,000 per annum.

A point of interest which never seems to have been investigated, except at W.O.S.B.s, is the *reliability* of vocational guidance or selection procedures as a whole. If the same group was tested and interviewed by two psychologists how closely would their recommendations agree? However, reliability is of little importance provided that validity is adequate.

Validation of War Office Selection Board Procedures

Fitts (1946) points out that there is no scientifically acceptable evidence whatever as to the value of the German officer selection procedures. A great deal of work was done in this field by the British Army Research and Training Centre, and this will be reviewed here since it was concerned chiefly with the value of the selection as a whole rather than with that of the separate tests or separate board members. If its results are somewhat disappointing, this may be attributed largely to the following difficulties:

(1) Owing to the lack of technical control, different boards used different methods and had different standards. It was hardly feasible, therefore, to treat candidates graded high, medium or borderline from different boards as equivalent. Again the methods and the recording systems changed at intervals during the war. Pass rates also varied markedly with operational requirements.

(2) The "selectivity" of the population was so high that even highly valid procedures would be expected to achieve only moderate or low coefficients. Only men regarded as promising by P.S.O.s or by their C.O.s reached W.O.S.B.s at all, that is barely 10 per cent. of the Army population. Some 60 or 80 per cent. of these candidates were rejected by the boards, and it was, of course, impossible for rejects to be sent forward in order to prove what bad officers they would make. Another 5 to 20 per cent. of accepted cadets failed in their O.C.T.U. training, and during subsequent field training an unknown proportion of the less suitable got shunted off into relatively unimportant jobs. Thus, only the final cream, amounting perhaps, to 50 per cent. of W.O.S.B.

acceptances or about 15 per cent. of W.O.S.B. candidates reached front line duties.

(3) Men passed by any one board often go to different O.C.T.U.s where standards differ and where, very often, the instructors who know them best are shifted away to new jobs before the follow-up investigator can catch them. Thus it is impossible to collect any large group who have both been selected, trained and assessed uniformly. At the operational stage this scattering is, naturally, even more marked. Only one or two cases, probably passed by different W.O.S.B.s and trained by different O.C.T.U.s at different periods, are likely to be found in any one unit.

(4) While information is rather more easily collected from O.C.T.U.s than from the field, it is extremely doubtful whether a cadet's O.C.T.U. grades correspond closely with his ultimate value as an officer. An American investigation (cf. Jenkins, 1947) gave a correlation of only + ·15 ± ·075 between officer school and subsequent combat ratings*. A follow-up of 329 officers in various British units, four to thirteen months after commissioning, i.e. before they had been in action, yielded correlations averaging + ·26 between O.C.T.U. passing-out grades and C.O.s' opinions.

(5) Although a technique was eventually evolved for securing reasonably reliable judgments from C.O.s it was found that their gradings were greatly affected by the subalterns' age, and the length of time they had been commissioned, and that they varied in different Arms. For accurate validation, therefore, it was desirable to secure large groups in which all these factors were held constant—a practically impossible task.

One successful comparison of old board and new board cadets was made possible by the fact that, during 1942, several boards of each type were running simultaneously before the old ones were finally superseded. Ratings were collected on some 1,200 cadets by follow-up officers who held conferences with the O.C.T.U. instructors (who did not know which cadets came from which board). The gradings were reduced to a three-point scale with the results shown in Table V.

---

* It is not clear how this figure was reached. From the published tabulations of Officer School leadership ratings and subsequent combat ratings, the present writers obtain a tetrachoric correlation of ·32.

TABLE V.—TRAINING RESULTS OF OFFICER CADETS SELECTED BY TRADITIONAL AND BY PSYCHOLOGICAL METHODS

Board	*No.*	*Above Average Grades*	*Average*	*Below Average and Fail*
		per cent.	*per cent.*	*per cent.*
Old . . .	491	22·1	41·3	36·6
New . . .	721	34·5	40·3	25·2

The selectees from seven out of the eight new boards were superior to those from all five of the old boards, and this superiority was found (though not always to a statistically significant extent) in all ten O.C.T.U.s, representing different Arms. Actually the result is less favourable than it looks. A true correlation cannot be calculated, but it is unlikely that the figures given in Table VIII correspond to a higher coefficient than about + ·2. Nevertheless, the combination of slightly more valid selection procedures with greater attractiveness to candidates resulted in the sending of two-and-a-half times as many above-average cadets to O.C.T.U.s as the old boards would have done, within five months after the establishment of new boards.

The follow-up of 329 officers mentioned above, some of whom were old and some new board products, yielded almost identical results at O.C.T.U. Any differences between the two groups on C.O.s' ratings in their units were too greatly masked by age, Arm, and length of commissioning to be calculable. A small-scale follow-up of thirty-six R.A.F. candidates for administrative commissions gave promising results. These men were both interviewed by a R.A.F. interview board and put through the W.O.S.B. procedure. Their grades at the end of eight weeks' O.C.T.U. are said to have correlated fairly closely with the W.O.S.B. predictions (how closely was not published), not at all with the interview board predictions.

Several later enquiries in the field showed that C.O.s were well satisfied on the whole with W.O.S.B. products, only some 7–13 per cent. being considered unsatisfactory either in this country, in the Mediterranean theatre, or in the Army of the Rhine. Thus, in spite of the continuous drainage of the best material, W.O.S.B.s appear to have maintained the flow into the sixth year of the war. But it was clear that C.O.s' standards were to some extent dependent on what they received, i.e. that their borderline of unsatisfactoriness may have dropped. As no control group of subalterns not sent through W.O.S.B. was available, these figures prove little. Similar results were obtained in an A.T.S. officer follow-up. A correlation of + ·28 (uncorrected for selectivity) was obtained between A.T.S., W.O.S.B. and O.C.T.U. gradings, but the agreement of W.O.S.B. with assessments of A.T.S. officers in their units, although positive, was very small.

In 1945, assessments of over 500 officers in Infantry and Royal Artillery were obtained from C.O.s just before the crossing of the Rhine, and slight but significant differences in efficiency were found between those who had been passed as As or Bs, Cs and Ds at one or another of sixteen W.O.S.B.s. For all groups combined the correlation between the two sets of gradings amounts to about + ·165, but if correction is made for selectivity the coefficient to be expected (had all W.O.S.B. *candidates* gone forward) rises to about + ·35. It was also found that W.O.S.B. predictions were distinctly better for younger than for older men. Among those who were boarded at 23 years or under the uncorrected correlation was ·23, and among those who were 28 or over it dropped to ·06. Clearly the W.O.S.B.s had attached insufficient weight to age and experience—qualities which C.O.s particularly valued. Considering the unreliability of the criterion and the variations between boards, these figures are fairly satisfactory.

In the post-war period, operational follow-up has been impossible. But a large number of investigations have proved that W.O.S.B. grades give fair predictions of success at O.C.T.U. Under properly controlled conditions, correlations consistently reach the + ·4 to + ·5 level, though they depend greatly on the skill of the particular board members, and on the thoroughness of the O.C.T.U. assessments.

A fundamental weakness of W.O.S.B. selection was its unreliability, due to the large part played by uncontrolled subjective judgment. In one experiment 116 candidates were put through two boards a fortnight apart, and the average agreement was represented by a tetrachoric correlation of only ·67. Of candidates passed by any one board, 21½ per cent. were rejected or deferred by another board. Evidence was collected as to the agreement between one pair of boards using closely comparable methods, and another pair using highly "individual" procedures, also on the correspondence between the gradings of different board members. This was not published, but it is believed to have shown higher reliability among the technical than the non-technical members.

In 1945, however, an elaborate experiment, where two teams composed of highly experienced staff observed or interviewed the same 125 candidates, showed quite high reliability. Moreover, the agreement was at least as good when all members worked independently as when there was collaboration within, or between, teams. Both M.T.O.s and psychologists observed the leaderless group and other practical tests. Psychiatrists interviewed independently, but presidents sat in on each others' interviews. Table VI shows the average agreement as to the general suitability of candidates in the first column. The coefficients for pairs of observers are high, for interviewers rather lower, and for whole teams the final coefficient of ·80 is reasonably good.

In addition to general suitability judgments, ratings were given on fourteen to eighteen carefully defined traits and the median reliability coefficients in the second column show a similar trend. The last column gives the average correlation between each type of member and the final gradings of the candidates obtained at a conference of both teams under an independent chairman. It should be noted that all the figures quoted are averages based on several members, and that the separate coefficients for particular members often showed wide variations. Unfortunately, it was not possible to follow up these candidates and so to validate the various sets of judgments.

The conclusion that appears to follow from this review is that W.O.S.B. methods, applied haphazardly according to the whims

TABLE VI.—RELIABILITY OF JUDGMENTS BY MEMBERS OF TWO W.O.S.B. TEAMS

	Mean Reliability Coefficients	*Median Agreement on Separate Traits*	*Correlation with Final Disposal*
M.T.O. with M.T.O. . .	·86	·77	·83
Psychologist with Psychologist .	·78	·69	·83
M.T.O. with Psychologist . .	·79	—	—
President with President . .	·65	·68	·75
Psychiatrist with Psychiatrist .	·65	·47	·71
President with Psychiatrist . .	·62	—	—
M.T.O. or Psychologist with President or Psychiatrist . .	·59	—	—
Team with Team . . .	·80	·68	·91

of the staff are only of slight value, but that when standard techniques are evolved and applied uniformly by trained and experienced personnel, a satisfactory reliability may be obtained. Only under such conditions can good validity be expected, and a great deal more, admittedly difficult, research is needed to prove the value of the contributions of the various parts of the procedure. A number of such investigations are now under way.

Note on the Meaning of Correlation Coefficients

It is very difficult to convey the practical meaning of a correlation coefficient, say of + ·5. It does not imply 50 per cent. agreement. Indeed, many psychologists would say that it represents an accuracy of prediction only 13½ per cent. better than pure chance. The rationale of the figure, 13½ per cent., is as follows. Suppose that we guessed the actual output on the job of an employee as lying at the average for all employees, our average error of prediction might be, say, 100 units. If, however, we predicted output from the man's selection test score, our error would still be as big as 86½ units, on the average, hence the reduction would only be 13½ per cent. It is more enlightening, however, to consider what proportions of good or poor men are selected by a test with a certain validity coefficient.

Take as an example some job which only one out of every five unselected men can perform satisfactorily. If an employer has 1,000 candidates to choose from, and takes 200 at random, only forty of his choices will be satisfactory. Whereas if he applies a test, or tests, with a validity correlation of + ·5, and chooses the 200 men with the best test scores, eighty-eight of them (or more than twice as many) will be satisfactory. This seems to represent a great improvement, but the following Table VII shows that, from the individual candidate's standpoint, the tests are not very successful. When selection was random 160 were wrongly rejected, since they could have managed the job, and 160 who could not do it were wrongly accepted. The total wastage is, therefore, 320 or 32 per cent. When the tests are used the total wastage falls to 224, that is a reduction by only 30 per cent. Table VIII similarly shows the agreement when the validity of the tests is very low, and quite high. Not until the validity correlation reaches + ·74 is the wastage halved.

TABLE VII.—NUMBERS OF SATISFACTORY AND UNSATISFACTORY SELECTEES

A. Selected at Random

	Satis-factory at Job	*Unsatis-factory at Job*	
Selected	40	160	200
Rejected	*160*	*640*	*800*
Total	200	800	1000

B. Selected by Test Battery

	Satis-factory at Job	*Unsatis-factory at Job*	
Passed tests	*88*	*112*	*200*
Failed tests	112	688	800
Total	200	800	1000

TABLE VIII.—WASTAGE WITH TESTS HAVING LOW AND HIGH VALIDITIES

A. Test Validity .. + ·25

	Satis-factory	*Unsatis-factory*	
Passed tests	64	136	200
Failed tests	136	664	800
Total	200	800	1000

B. Test Validity .. + ·74

	Satis-factory	*Unsatis-factory*	
Passed tests	120	80	200
Failed tests	80	720	800
Total	200	800	1000

Unfortunately these Tables would not be correct if the selection ratio was altered, unless it became 4/5 instead of 1/5. If 80 per cent. of men could do the job, Table VII A would become Table IX, and all the others would be similarly reversed.

TABLE IX.—WASTAGE WITH HIGH SELECTION RATIO

	Satis-factory	*Unsatis-factory*	
Selected	640	160	800
Rejected	160	40	200
Total	800	200	1000

But with a selection ration nearer to ½, the wastage would be greater, and the more it diverges from ½ the more favourable the improvement produced by using selection tests. This can readily be seen from Fig. 1, which gives a graphical presentation of a correlation of about + ·5. Each dot represents one man's test score and his subsequent proficiency. When the selection ratio is $\frac{1}{2}$, the quadrant A represents men correctly selected by the tests, the quadrant C men correctly rejected, and the wastage, shown by the numbers in B and D, is fairly high. But when the ratio is very

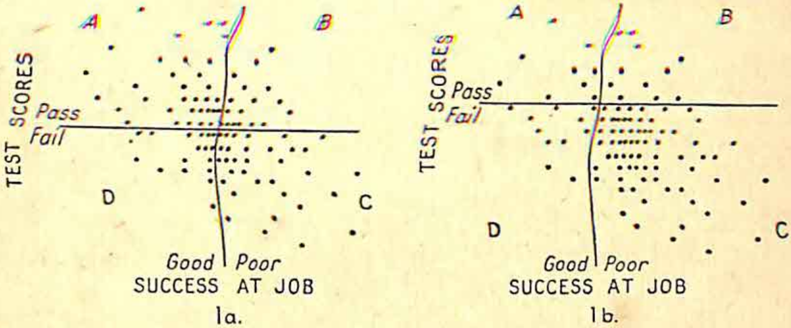


FIG. 1. SCATTERGRAMS OF CORRELATION COEFFICIENTS WITH SELECTION RATIOS OF 50 PER CENT. AND 20 PER CENT.

high or low as in Fig. 1b, it may be seen that the wrong acceptances, B, and wrong rejects, D, are relatively few.

Nevertheless, the *percentage reduction in wastage* does remain fairly constant so long as the selection ratio is not greater than 84 per cent. or smaller than 16 per cent., and this covers the great majority of vocational and educational situations. Fig. 2 is borrowed from McClelland (1942). Let the ratio be h per cent., then the wastage when no tests are used, or when their validity is represented by a correlation of zero, is $\frac{2h(100-h)}{100}$.

For example, if the ratio is 1 in 5 or 20 per cent., or alternatively 80 per cent., the wastage is $\frac{2 \times 80 \times 20}{100} = 32$ per cent., as shown in our Table VII A. For validities of ·25, ·50 and ·74 the graph shows that the number of mistakes amounts to 85 per cent., 70 per cent. and 50 per cent. respectively of this 32 per cent., so giving the numbers listed in Tables VII B and VIII. This same reduction would occur with other selection ratios in the middle range.

So far we have assumed that as many men are rejected or selected by the test as the proportion found unsatisfactory or satisfactory on the job. Naturally there is no need for this restriction. Although it is the commonest and most convenient plan, McClelland shows that, by an appropriate choice of pass mark on he tests, the wastage can be further reduced. In fact, for each election ratio and each validity coefficient, there is a certain test pass mark, readily determinable statistically, which produces minimum wastage. But to portray this situation would complicate matters, and the reader may be referred to the full Tables provided by Taylor and Russell (1939), and Tiffin (1946).

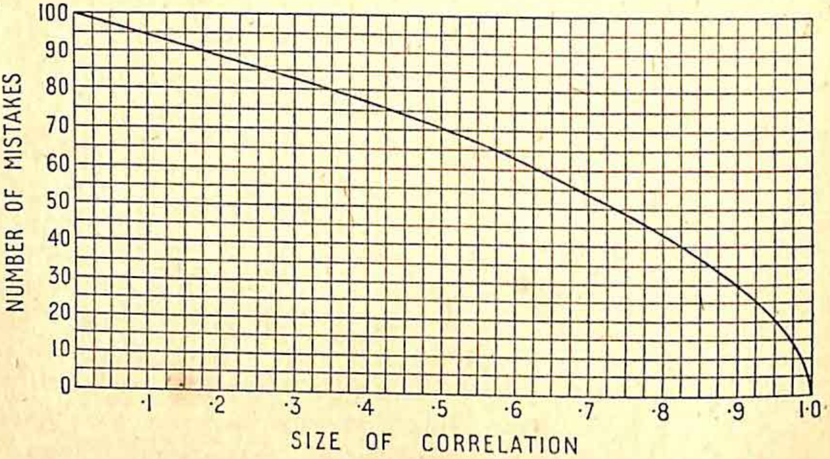


FIG. 2. PERCENTAGE OF WASTAGE REMAINING WITH DIFFERENT SIZES OF CORRELATION.

CHAPTER VIII

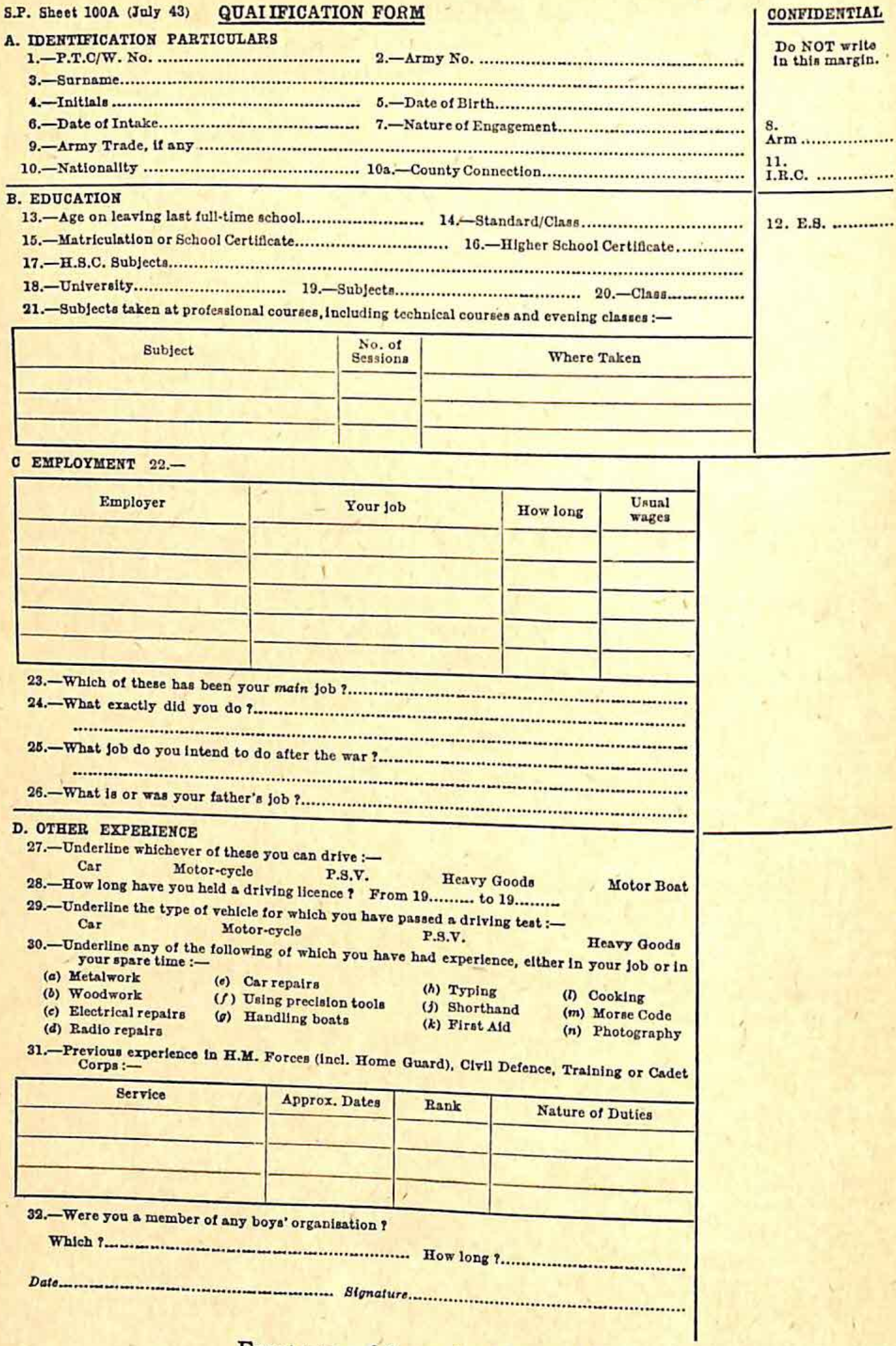
THE BIOGRAPHICAL QUESTIONNAIRE

Abstract.—A description is given of the Army qualification form and the naval biographical questionnaire, and attention is drawn to precautions in the construction and use of such documents.

Occupational experience, education, age, interests and other items recorded on questionnaires are shown to be significantly related to success in a variety of Service jobs. Examples are drawn from investigations of telegraphists, motor drivers, seamen, radio mechanics, etc. But there are considerable technical difficulties in turning such information into an objective form in order to yield direct predictions of occupational suitability. Single questionnaire items are very unreliable, especially among low-grade recruits, and they overlap in a very complex manner with one another and with intelligence.

Typical of the questionnaires used in the Forces is the Army qualification form, reproduced on pages 132 and 133. As may be seen it covers Service particulars, previous education and employment, experience that might be relevant to Army jobs, results of psychological tests and medical examination, and the P.S.O.s' conclusions. The questionnaire used in the Navy had additional sections on school and leisure time interests and on leadership experience (captain of a team, secretary of a club, being in charge of people, membership of pre-Service organisations, etc.). Though the questions are printed, they were read out by the tester to groups of recruits and further explanations were given when necessary. Incidentally, the front pages of these forms served as a useful and adequate, although unstandardised, test of literacy. In the case of the 1–2 per cent. of illiterates who could not produce an intelligible set of answers even under supervisions, the tester in charge would fill it in for them.

Now even the most straightforward questions may be misread or misinterpreted, especially by duller recruits. The following list of precautions is based partly on experience and experiments in the



S.P. Sheet 100A (July 43) **QUALIFICATION FORM** **CONFIDENTIAL**

Do NOT write in this margin.

A. IDENTIFICATION PARTICULARS

1.—P.T.C/W. No. .... 2.—Army No. ....

3.—Surname....................

4.—Initials .... 5.—Date of Birth....................

6.—Date of Intake.................... 7.—Nature of Engagement....................

9.—Army Trade, if any ....

10.—Nationality .... 10a.—County Connection....................

8. Arm ....

11. I.R.C. ....

B. EDUCATION

13.—Age on leaving last full-time school.................... 14.—Standard/Class....................

15.—Matriculation or School Certificate.................... 16.—Higher School Certificate....................

17.—H.S.C. Subjects....................

18.—University.................... 19.—Subjects.................... 20.—Class....................

21.—Subjects taken at professional courses, including technical courses and evening classes :—

Subject	No. of Sessions	Where Taken

12. E.S. ....

C EMPLOYMENT 22.—

Employer	Your job	How long	Usual wages

23.—Which of these has been your *main* job ?....................

24.—What exactly did you do ?....................

....

25.—What job do you intend to do after the war ?....................

....

26.—What is or was your father's job ?....................

D. OTHER EXPERIENCE

27.—Underline whichever of these you can drive :—

Car Motor-cycle P.S.V. Heavy Goods Motor Boat

28.—How long have you held a driving licence ? From 19......... to 19.........

29.—Underline the type of vehicle for which you have passed a driving test :—

Car Motor-cycle P.S.V. Heavy Goods

30.—Underline any of the following of which you have had experience, either in your job or in your spare time :—

- (*a*) Metalwork
- (*b*) Woodwork
- (*c*) Electrical repairs
- (*d*) Radio repairs
- (*e*) Car repairs
- (*f*) Using precision tools
- (*g*) Handling boats
- (*h*) Typing
- (*j*) Shorthand
- (*k*) First Aid
- (*l*) Cooking
- (*m*) Morse Code
- (*n*) Photography

31.—Previous experience in H.M. Forces (incl. Home Guard), Civil Defence, Training or Cadet Corps :—

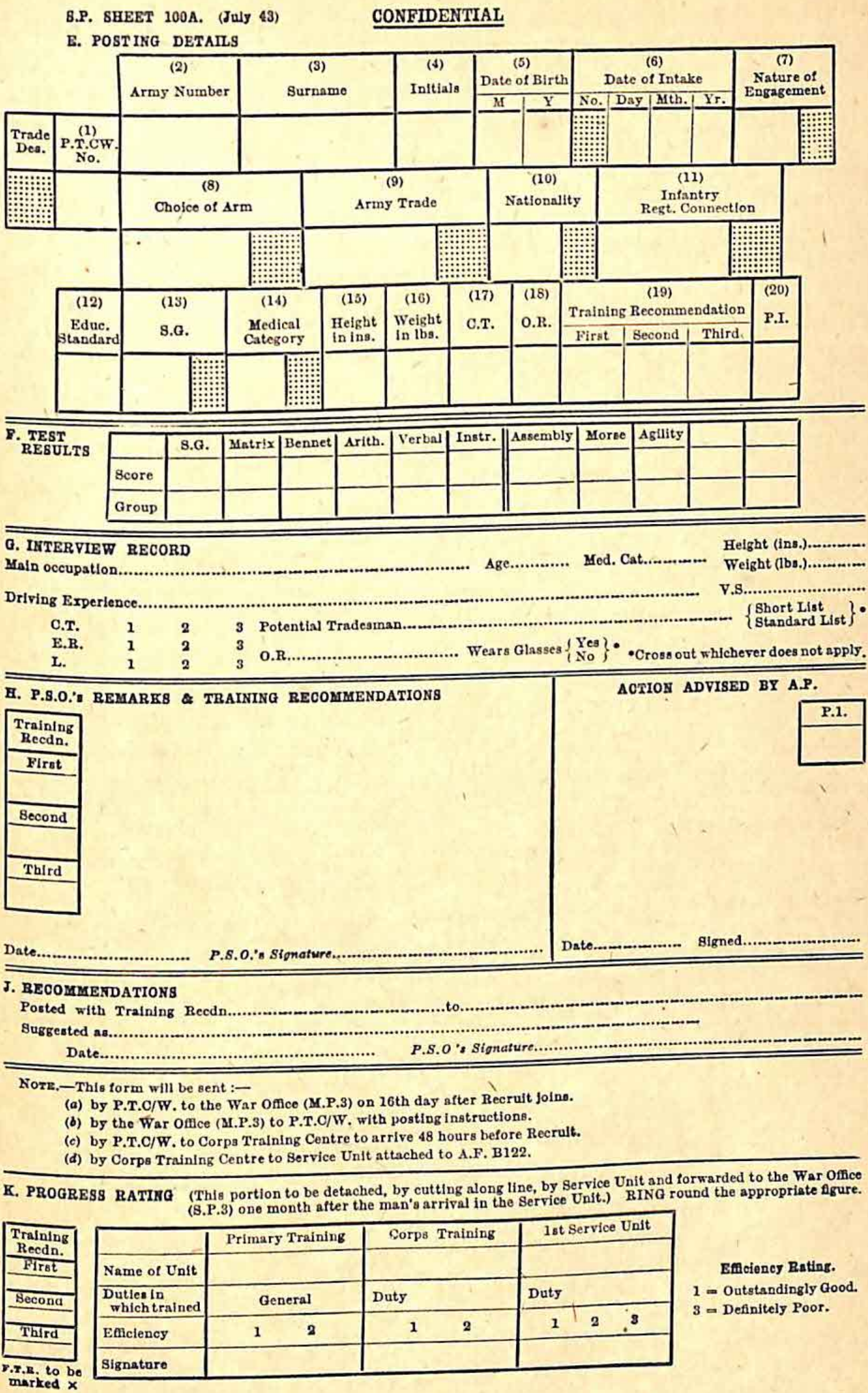
Service	Approx. Dates	Rank	Nature of Duties

32.—Were you a member of any boys' organisation ?

Which ?.................... How long ?....................

Date.................... *Signature*....................

Front page of Army Qualification Form.



S.P. SHEET 100A. (July 43) **CONFIDENTIAL**

E. POSTING DETAILS

Trade Des.	(1) P.T.CW. No.	(2) Army Number	(3) Surname	(4) Initials	(5) Date of Birth		(6) Date of Intake				(7) Nature of Engagement
					M	Y	No.	Day	Mth.	Yr.	

(8) Choice of Arm	(9) Army Trade	(10) Nationality	(11) Infantry Regt. Connection

(12) Educ. Standard	(13) S.G.	(14) Medical Category	(15) Height in ins.	(16) Weight in lbs.	(17) C.T.	(18) O.R.	(19) Training Recommendation			(20) P.I.
							First	Second	Third	

F. TEST RESULTS

	S.G.	Matrix	Bennet	Arith.	Verbal	Instr.	Assembly	Morse	Agility		
Score											
Group											

G. INTERVIEW RECORD

Height (ins.)..........

Main occupation.. Age........... Med. Cat........... Weight (lbs.)..........

Driving Experience.. V.S.....................

C.T. 1 2 3 Potential Tradesman.................................... {Short List / Standard List}*

E.R. 1 2 3

L. 1 2 3 O.R.............................. Wears Glasses {Yes / No}* *Cross out whichever does not apply.

H. P.S.O.'s REMARKS & TRAINING RECOMMENDATIONS

Training Recdn.
First
Second
Third

Date.......................... *P.S.O.'s Signature*..................................

ACTION ADVISED BY A.P.

P.I.

Date............... Signed.....................

J. RECOMMENDATIONS

Posted with Training Recdn...............................to..

Suggested as...

Date.. *P.S.O 's Signature*..

NOTE.—This form will be sent :—

(*a*) by P.T.C/W. to the War Office (M.P.3) on 16th day after Recruit joins.

(*b*) by the War Office (M.P.3) to P.T.C/W. with posting instructions.

(*c*) by P.T.C/W. to Corps Training Centre to arrive 48 hours before Recruit.

(*d*) by Corps Training Centre to Service Unit attached to A.F. B122.

K. PROGRESS RATING (This portion to be detached, by cutting along line, by Service Unit and forwarded to the War Office (S.P.3) one month after the man's arrival in the Service Unit.) RING round the appropriate figure.

Training Recdn.
First
Second
Third

F.T.R. to be marked ×

	Primary Training	Corps Training	1st Service Unit
Name of Unit			
Duties in which trained	General	Duty	Duty
Efficiency	1 2	1 2	1 2 3
Signature			

Efficiency Rating.

1 = Outstandingly Good.

3 = Definitely Poor.

Back page of Army Qualification Form.

Forces, partly on the large published literature dealing with the questionnaire (cf. Koos, 1928; Symonds, 1931; Goodenough and Anderson, 1931; Good, Barr and Scates, 1936; Vernon, 1938a, 1939b):

(1) The design and layout of the questionnaire should make it easy to follow. Bad design may lead to omissions of questions and to cramping of the written responses.

(2) The instructions should be clear and simple, couched in words of a suitable vocabulary level. Questions should be unambiguous and, as far as possible, concrete and specific. A preliminary trial with a moderate-sized group of informants is extremely useful for showing up obscurities.

(3) If selective or choice-response questions are to be included, a trial with open questions will reveal responses which might not have occurred to the psychologist. Too few alternative answers are obviously undesirable, but too many are liable to confuse. Actually, owing to the danger of misunderstandings, it is usually better to avoid choice-response questions, for example, listing various types of post-primary schooling and asking recruits to tick the ones that apply to them. Such a form is certainly easier to code on to punched cards, or to subject to statistical analysis, but it is always liable to blur or distort the almost infinite variety of human histories. When investigations are made, say, into the predictive value of previous education, the diverse responses can readily be classified by the psychologist. The only background questionnaire in the Forces which was coded was that used by the W.O.S.B.s, where the informants were, of course, above average in literacy.

(4) Items dealing with interests or types of vocational experience are very liable to distortion, owing to varying standards of judgment. Some informants will tick almost every item, whereas others who may, in fact, have spent as much or more time on the activities listed, only tick a few to which they are particularly attracted. The degree of interest or amount of experience may be checked in a subsequent interview, but standardised tests are preferable if time allows (cf. p. 137).

(5) Questions which appear to involve "right" or "wrong" answers should be avoided in order to reduce any inclination

to falsify. Again, emotional terms or prestige names which may touch off prejudices or stereotypes are undesirable. If questionnaire tests or inventories dealing with attitudes or personal traits are to be applied (cf. Chapter XV), they should be kept distinct from this essentially factual enquiry. In general the motivation of the informants requires careful consideration. There should, in other words, be strong reasons why they should take the trouble to fill up the questionnaire accurately.

(6) It should be borne in mind that the questionnaire is in essence an interview, to which answers are given in writing, and that it requires the same skilful formulation of questions, the same tact, as does an interview. While some informants are suspicious of this impersonal instrument and reveal themselves more willingly in the face-to-face interview situation, others are "bad interviewees" who may be more frank in a questionnaire.

Validity of Biographical Items

The objectively scored application blank has long been a popular technique among American industrial psychologists. The relevance of each item is determined empirically and a scoring key prepared for use with future applicants. This method was extended to air crew in the Royal Canadian Air Force, to U.S. Army officers and other groups during the war, but was never applied systematically in this country. Nevertheless, the relation of questionnaire items to occupational success was worked out in some seventeen branches of the Royal Navy and Army, and the predictive value of particular items such as occupation, educational status, experience of driving or morse, age, etc., in a great many more jobs. It is not possible to summarise all these results, but some specimen investigations will be described below. First, however, certain difficulties should be pointed out which are just as likely to arise in industrial or educational practice as in the Forces.

(1) The validation of an item to which only two responses are possible requires much larger numbers of cases than the validation of a test or other extended variable. Roughly double as many are needed to give equally trustworthy correlations. Thus an investigation with less than about 400 cases may be misleading since its indications of the relevance

of certain items might not be borne out in another similar group. This fact renders the study of previous occupation particularly difficult since only quite small numbers are likely to have belonged to any one occupation. Occupations must, therefore, be classified into very broad categories, and in the absence of any generally accepted scheme, such classification tends to be subjective and inconsistent. Significant differences between the proficiency of different occupational groups cannot be expected unless each group contains more than, say, 40 cases; but such groups may be attainable only if several heterogeneous occupations are combined. For instance, "clerks" might appear to be a straightforward group. Yet, in a study of naval officer cadets, accountants, Civil Servants, solicitors' and Local Government clerks were excellent, while works clerks, costing, railway and shipping clerks were below average, and bank clerks and insurance clerks were intermediate (an average of 75 in each group). The name of the occupation may also be a very poor guide, especially with 18-year-olds. Thus, mate may mean anything from a highly skilled apprentice to a low-grade labourer.

(2) Items which show fair correlations with proficiency may do so merely because they involve intelligence, or because they overlap with other items. Taking occupation again, clerks were found to be superior in numerous naval branches, including some of the mechanical ones, but this was due almost entirely to their high intellectual level which made them excellent trainees for any job. When intelligence was held constant they were found, for example, to be poorer in active and fighting qualities in Coastal Forces, and to provide an excess of failures in certain mechanical jobs. Similarly the good results of those who preferred science or mathematics at school to geography, handwork, etc., mainly reflects the better schooling and intelligence of the former group. Even the advantages of membership of pre-Service organisations (A.T.C., Scouts, etc.) was found to be largely due to the same factors. Thus, an item analysis of a questionnaire should at least determine the relation of each item to intelligence as well as to occupational proficiency, and partial correlation (or analysis of covariance) should be used to

indicate the independent contributions of each item*.

(3) As already mentioned, reliance cannot be placed on item scores obtained from a single group. The scoring should be applied to a fresh group and its validity determined.

(4) The answers given to many items are so unreliable that it is better to substitute tests whenever possible. Hence oral or written tests of trade knowledge were introduced, also practical achievement tests in morse receiving, shorthand and typewriting. It was noted in the Navy that much more exaggerated claims to experience of morse were made at recruiting centres, where there was no likelihood of their being put to the test, than in H.M.S. *Royal Arthur*, where allocation to signaller or telegraphist branches was imminent. That genuine experience was, if anything, more valuable than good intelligence test scores was proved among telegraphists, air gunners and other "communications" ratings. Nevertheless, examples were found of men claiming to receive at 8 words/min. who failed their morse training, and of men claiming 15 words/min. who required to be backclassed before they achieved the requisite accuracy at 20 words/min. Experience of 4 or fewer words/min. was found to be practically worthless, since morse is often so badly taught in cadet units as to handicap the early stages of naval morse training. It seems highly probable that carefully designed, even if brief, tests of interests would similarly have more diagnostic worth than merely ticking each of the activities in which recruits think themselves interested.

(5) Selectivity interferes with the significance of questionnaire items in even more complicated fashion than it does with test scores. For example, when P.S.O.s have creamed the really experienced fitters and turners for the high-grade mechanical branches, the follow-up of occupation in a less-skilled mechanical branch is likely to show plumbers and sheet metal workers as superior to fitters and turners. Another example may be drawn from communications, where it was found on more than one occasion that men claiming mechanical interests were actually poorer than average as signallers or telegraphists. This may well have

---

* Preferably multiple correlation, discriminant function, or multiple contingency methods should be used in order to allow for the overlapping of items with one another and with test scores.

been due to the drafting of more responsible and more able mechanically-minded recruits to other branches. Theoretically it might be possible to correct for such selectivity, but no simple statistical technique is as yet available for analysing selectivity in a large number of yes-no items.

(6) Items other than two-response ones frequently show non-linear relationships to proficiency, so that ordinary correlation techniques are inappropriate. For example, neither recruits possessing 2–4 words/min. morse experience, nor those with 1–6 months' driving experience, were found superior to men with no morse or driving experience, but greater degrees of experience did give an advantage. Among naval officer candidates those leaving school at 16–17 were superior to those leaving at 14–15, but those educated to 18–20 showed no further rise. Age tends to show very irregular effects. In some branches 20–25 year olds were superior both to 18–19 year and to older men. In communications age *below* 20 made for more rapid morse learning, but thereafter there was apparently no big drop till after 30 or 35 years. In Army motor drivers age up to 25 was irrelevant, but then (at least up to 40) the number of hours of training needed was roughly identical with years of age. The significance of such variables must, therefore, be explored by techniques like chi-squared or correlation ratio, which makes it difficult to combine them, or to determine their overlapping, with other items.

In spite of these difficulties a number of very substantial associations were discovered between questionnaire items and proficiency. In the Navy it was particularly noticeable that leadership experience was advantageous, not only among seamen and officer candidates, but in diverse employments such as air fitters, telegraphists, radio mechanics and safety equipment ratings. Scouts, members of Boys' Brigade, and the like, and of organisations such as the A.T.C. and Sea Cadets showed similar superiority, though the significance of this finding is—as pointed out—dubious.

Table X shows some specimen correlations between occupational experience and proficiency in the A.T.S., also between the best weighted battery of selection tests and proficiency (uncorrected for selectivity). Although so troublesome to assess, occupational success is clearly more predictive than test ability in several jobs. This is particularly true in war-time when training has to be as short as possible.

TABLE X.—PREVIOUS OCCUPATION, TESTS RESULT AND PROFICIENCY IN THE A.T.S.

Civilian Experience	*A.T.S. Employment*	*Correlations with Proficiency*	
		Experience	*Test Battery*
Clerical	General Duties Clerks — Overall proficiency	·296	·673
	General Duties Clerks — Stenographic proficiency	·671	·371
Clerical	Teleprinter Operators	·487	·523
Domestic work	Cooks	·204	·390
Switchboard operating	Switchboard Operators	·533	·485
Motor driving	Drivers	·510	·375

The collected results on driving experience from seven representative driver training regiments in R.A. or infantry are shown in Table XI. The final driving proficiency marks or grades are classified into good (the best 30 per cent.), medium (42 per cent.), and poor (the bottom 28 per cent., including failures). Apparently amount of experience beyond seven months has little influence. If a tetrachoric correlation is calculated the coefficient of ·56 is similar to the A.T.S. result, and exceeds the multiple correlation of ·42 obtained between the standard test battery and proficiency. Since experience showed negligible (positive) correlations with intelligence, a combination of experience and test scores should theoretically yield much improved predictions. Actually, owing to the shortage, experienced drivers were accepted regardless of intelligence levels, and minimum test standards applied only to the inexperienced.

TABLE XI.—EFFECTS OF PREVIOUS EXPERIENCE ON DRIVING PROFICIENCY

Driving Experience before Recruitment	*N*	*Proficiency*		
		Good	*Medium*	*Poor*
		per cent.	*per cent.*	*per cent.*
Professional	75	48	44	8
Amateur for more than 2 years	55	47	40	13
7 months to 2 years	58	47	41	12
6 months or less, slight or casual	43	19	42	40
Nil	180	15	42	43
Total	411	30	42	28

Two rather extensive naval investigations are worth summarising in some detail. Among 1,000 seamen and stokers undergoing elementary training, instructors picked out the best three and worst three from their own classes, yielding 136 "bests" and 136 "worsts" in all. S.P. tests 1–4 all gave very similar correlations, the multiple coefficient being ·444 and the correlation of T2 ·433. Items are listed below in order of their statistical significance (T2 being held constant by analysis of covariance).

Age. 35 per cent. of men over 30 and only 9 per cent. of men under 20 were bests; the relationship was approximately linear.

Occupation. Retail tradesmen and drivers *v.* the rest.

Leadership experience.

Membership of Scouts or other boys' organisations.

Labouring occupations *v.* the rest. Here the association was negative.

"Social" interests. Ticking the items playing music, concert party work, acting, motor driving.

Clerical occupations *v.* the rest. This showed the highest direct relationship to proficiency, but it was mostly attributable to high intelligence.

Service experience, in Home Guard, J.T.C., A.T.C., Civil Defence, etc.

Education beyond 14, or—negatively—left school at 14 in a class below the top.

Athletic interests. Ticking the items football, boxing, athletics, swimming or camping.

Domestic interests. Ticking the items reading, house repairs, gardening, cooking.

Emotional instability, a "P" rating gave a negative association. Only 43 men were so marked, but 23 per cent. of these were worsts as contrasted with 13 per cent. of non-"Ps." This item, together with mechanical interests and job experience were not significant statistically.

The relationship with proficiency and the inter-item relations were expressed as correlations. Multiple correlation showed that the most predictive items in combination were T2, age, occupation, leadership, scouting, Service experience, emotional instability, domestic interests, and education. A rough scoring scheme for items other than T2 was made out. When the questionnaires were scored in this way a biserial correlation of ·461 was obtained, and for T2 + questionnaire score combined, the correlation was ·533. No opportunity occurred for verifying these figures on a fresh group.

One other illustration is drawn from a very high-grade naval branch, radio mechanics, for which selection was particularly difficult owing to its extreme specialisation. Groups of 850, 500 and 300 were followed up at different times and the correlations of T2 with course results were only ·338, ·209 and ·286 respectively. Multiple correlations based on Tests 2 and 3b were only slightly higher, but in the last two groups an advanced mathematics paper and an electrical information test gave correlations of ·332 and ·457. It was noted that all men passing their training were very high test scorers, but that by no means all high scorers passed; also that the validity of S.P. tests varied significantly in the different centres where the mechanics were trained. These observations suggest that morale and willingness to work at so long and complex a course play a very large part. These are factors which may vary considerably in different centres, but which may also depend on the motivation of individual students. Hence it might be possible to measure them by means of a questionnaire. An educational and interests questionnaire was given to the group of 500, and the following items were found to be significantly associated with good course results:

Staying on at school after the age of 15.

Attendance at continuation or evening classes (not necessarily technical; banking, insurance and accountancy students were the best, presumably because they showed greatest "seriousness of purpose").

Having passed any examination.

Obtaining credit or distinction in School Certificate or Matriculation *v.* passing or merely reaching "School Certificate standard."

School Certificate or Matriculation in mathematics and/or physics.

Ticking interests in metalwork, house repairs, radio repairs, electrical repairs, or photography.

Among the passes 43 per cent. and failures 7 per cent. showed all five favourable educational signs, while 10 per cent. and 33 per cent. respectively showed only 2, 1 or 0 signs, corresponding to a correlation of ·42. Again, 56 per cent. and 29 per cent. respectively checked two or more of these interests, corresponding to a correlation of ·22. These two factors and T2 were almost independent, and a scoring scheme based on all three gave the very promising multiple correlation of ·56.

CHAPTER IX

THE INTERVIEW AND ITS VALIDITY

Abstract.—Hints are given on the conduct of employment or diagnostic interviews, with special reference to the contributions of Oldfield, Wilson and others. A number of the points are listed into which Service psychologists, psychiatrists and P.S.O.s enquired when attempting to assess the attitudes and emotional stability of recruits. Previous research shows considerable inconsistency between the judgments of different interviewers of the same man, and suggests that such judgments are often so biased as to have very poor validity. While it is admitted that the personality as a whole can only be summed up by the "clinical" approach of an interviewer, there is evidence that the "psychometric" approach —which measures partial facets of the person—may actually give better predictions of occupational suitability.

Investigations in the Services tended to confirm the unreliability of interviewing, and to show that test results alone have better validity than the average P.S.O. interviewer. But there were large individual differences in the skill of different P.S.O.s, and the value of interview judgments by qualified psychologists and psychiatrists was generally high. It is pointed out that, in spite of its defects, interviewing is often more economical than an elaborate psychometric programme, and more acceptable to the persons undergoing classification and to their employers.

The technique of interviewing, and the value of conclusions based on interviews, have been widely studied by psychologists. Useful references include Anderson (1929), Bingham and Moore (1931), Symonds (1931), Strang (1937), Rodger, Nadel and Brown (1939). We shall summarise some of the main points briefly, and then consider in more detail recent contributions by British writers.

(1) The large variety of types of interview, or of aims of interviewing, should be recognised, since they require different approaches and techniques; for example, the diagnostic

interview (vocational or psychiatric), the treatment interview (psychotherapeutic or the giving of vocational or educational advice), the research interview for collecting information (the social worker's, anthropologist's, public opinion surveyor's, and others). Interviewing by a board is generally considered more artificial and more disturbing than the ordinary two-person interview.

(2) Though mechanical rules cannot be laid down, the technique of interviewing can be formulated and greatly improved by training. Interviewers should also be carefully selected.

(3) The main qualities of a good interviewer, and the main factors leading to good rapport, are thorough knowledge of the jobs or other matters with which the interviewee is concerned and of topics in which he is interested, emotional maturity or a well-adjusted personality such that the interviewer is not shocked by anything the interviewee says, reputation among previous interviewees for sincerity, sympathy and sensitiveness. Good health and freedom from fatigue or strain are also valuable.

(4) While many writers urge greater standardisation of procedure, and while an oral questionnaire may be adequate for some purposes, e.g. public opinion research, most psychologists, in this country at least, insist on flexibility. They may have a scheme of points to cover, or hold a general plan in mind, but they prefer not to keep to any fixed order of topics since this makes for awkward switches. The interview is not an experimental situation; it is more effectively controlled and standardised by aiming at good rapport than by using stereotyped questions.

(5) All available information should be ready, and should be critically studied, beforehand; for example, test scores and material which can appropriately be collected by questionnaire.

(6) Nervousness can be reduced by appropriate treatment of the waiting interviewee, and by suitable arrangements of the physical environment in which the interview is held. Equally important is a calm, unhurried approach, giving the impression of plenty of time to spare.

(7) Opening enquiries should be factual or otherwise designed

to put at ease. Topics likely to arouse shyness, guilt or inferiority should generally be approached indirectly. But the course of the interview must be adapted to the interviewee's personality and may be very different with different people.

(8) In general the interviewee should do most of the talking; at least there should be provision for spontaneity and initiative on his part. Information obtained by interrogation is known to be less reliable than spontaneous narrative.

(9) A useful way of getting rational and impersonal information on traits and interests is to invite comparisons: "Do you prefer doing X to Y? Are you like your brother (or sister) in Z?"

(10) The taking of notes depends on the interviewee and the topic under discussion. Notes should be legible and concise. It is of great importance to have records of factual evidence to back up any conclusions reached or recommendations made, and not to trust to memory. Information received and observations made should be carefully distinguished from interpretations.

(11) There is ample evidence that any prejudice on the part of the interviewer may unwittingly affect the information he collects, or judgments he makes. In diagnostic interviewing he should particularly try to avoid being biased by one or two outstanding merits or defects in the interviewee, and by his first impression of him, or by chance resemblance to acquaintances, thus neglecting many of the complexities of personality.

(12) The belief in external signs of personal traits (particular features, type of build, gestures, speech, dress, etc.) is highly delusive. Manner and expressive movements may, indeed, be significant, but must be interpreted with caution since they may arise from different causes in different people, and they are largely dependent both on transient moods, e.g. temporary nervousness, and on cultural conventions. There is probably no relation whatever between any static facial or bodily characteristic and any vocational skill (apart from crippling abnormalities).

(13) In the diagnostic interview it is generally useful for the interviewer to express his main conclusions in the form of ratings

on a small number of important and more or less independent personality traits. Such quantification helps to produce a thoughtful analysis of the evidence, and is valuable in follow-up. The ratings may be supplemented by a pen-picture of the person as a whole.

(14) When an interview involves guidance, the interviewer should sum up the decisions reached at the close, and should ensure that the interviewee is set to take some appropriate action.

Recent Discussions of Interviewing

Among publications since 1939, Oldfield's book (1941) is outstanding for its analysis of the steps by which interviewers reach their conclusions. In his view the aim of the employment, and many other types of, interview is not to collect factual information nor to make judgments from this information or other clues, but to stimulate the interviewee to display his "attitudes," and to observe these attitudes. By attitudes he means consistent ways of reacting to situations or topics, or consistent types of behaviour, which are more specific and more overt than the underlying personality traits. Such attitudes are partly inferred from expressive movements (manner, speech, etc.) and partly from information about past history, but chiefly by raising various topics and seeing how he reacts, by observing the effects on him of sympathy, humour, agreement, disagreement, interest, surprise, etc. We thus get "the growth and continuously increased articulation of a general schematic picture of his (the interviewee's) personality in the interviewer's mind." This *schema* or working model can then be imagined as fitting, or not fitting, into the prospective job. Generally, however, it is implicit or unformulated and the psychologically untrained interviewer has considerable difficulty in putting it into words. He tends to subsume it under one or another of a set of stock types or frameworks which inevitably distort the complex personality structure.

Wilson (1945) has described the methods he employs in interviewing candidates for high-grade employment or training, such as officer cadets, skilled mechanics and engineers. His main theme is that an interview which can last only about twenty minutes should be a strenuous interplay between the two persons concerned, not just a pleasant conversation leading to a number of intuitions. No questions should be so superficial that they can be answered merely by yes or no. The interviewer should consciously control and vary the "pace" of the interview. He should possess a clear idea of his objectives and should try to collect evidence under a set of headings such as those listed below, in terms of which Oldfield's *schema* may be formulated and the suitability of different candidates compared. These do not constitute topics which are taken up *seriatim*; rather the course of an interview consists of an oral autobiography, prodded by the interviewer, and any items may contribute data under several headings.

(1) *Effective Intelligence.*—What use has the candidate made of the ability shown by his test scores; what are the quality, pace and efficiency of his thinking and self-expression? Can he analyse, abstract and generalise about novel and complex topics, or is his thinking merely reproductive, vague or muddled? Does he show intellectual assurance, clarity and fluency?

(2) *Technical Prowess.*—Just as effective intelligence differs from test scores, so here the interviewer tries to go beyond the candidate's academic or industrial record. What use can he make of his book-work? Does he understand basic ideas and can he apply them to new practical problems resourcefully? Is his technical capacity confident, dilettante, original, or restricted?

(3) *Interests.*—Their number, range and depth, their stability or changeability, whether integrated or conflicting, and whether relevant to the proposed employment. The interviewer should conjure up the atmosphere of the job to see whether it may be distasteful. How did the main interests arise; are they autonomous or based on some temporary identification?

(4) *Motivation.*—Soundness of reasons for desiring employment (or a commission), and strength of drives for undergoing long and tiresome training. Was his motivation steady in the past, showing responsibility, pride in achievement, desire to complete any task undertaken? Does he work to inner standards rather than outer demands? Has he a useful degree of compulsive or obsessional tendency? Does he possess a detailed and realistic knowledge of the proposed job or training course, and has he taken the necessary preliminary steps?

With this are linked his energy or forcefulness and his doggedness or persistence in the past, viewed in the light of his vocational opportunities.

(5) *Dominance* or competitiveness, i.e. his managerial and leadership potentialities.

(6) *Acceptability* to people he will work with, under or over, including social and athletic accomplishments, freedom from annoying habits, conventionality of outlook, liking for company. This is of importance even among such persons as skilled mechanics whose work appears to involve few social contacts.

(7) *Personal Attitudes.*—Whether co-operative or individualistic, irritable or impatient. Who is he influenced by or dependent on, e.g. what are his relations to his parents? What are his reactions to the Service and authority, also to exerting authority?

Numerous hints or clues are listed by Wilson which may throw light on these headings, or which should be distrusted. Most reliance should be placed on the candidate's past record (as amplified orally) and on his test results. One feature of Wilson's system that appears a little dangerous is his advocacy of informal achievement, aptitude and personality *tests* during the interview; for example, setting problems to bring out mental or technical ingenuity and pertinacity, observing whether he "crumbles" when the "pace" of the interview is too great, or whether his "dominance" comes out in his attitude to the interviewer. There is so much evidence of the lack of validity, and of divergences in interpreting, such uncontrolled tests that interviewers should not generally be encouraged to use them. Valuable factors, which he also stresses, are the possession by the interviewer of sufficient information about any and every technical, cultural and athletic field to be able to probe the thoroughness of the candidate's relevant interests, of adequate psychological knowledge to recognise abnormal motivations, and of familiarity with the "mythology" which candidates such as recruits hold regarding the advantages and disadvantages of various employments. Finally, while Wilson prefers separate interviews by the psychologist, the medical officer and the technical examiner, etc., he considers that candidates should not be accepted or rejected on the basis of any one of them, but should be discussed at a conference where the findings can be integrated, and they can be considered from all angles.

An article by Misselbrook (1946) gives a useful picture of the interview at, as it were, a lower level, that is the P.S.O.s' ten to fifteen minute interview with candidates for the great bulk of Service employments where the effective use of the available time is of the utmost importance. He points out the value of the preliminary talk to each batch of recruits, of the information charts (the extent to which a recruit has studied these gives a useful index of the seriousness of his attitude to an employment), and particularly of the questionnaires as checked up by the recruiting assistants. If the interviewer is practised in "sight reading" such questionnaires, he or she can adapt a very flexible approach, since many of the essential points are already covered. Among other aids to the harassed interviewer are sets of oral trade questions which, if known by heart, can be applied informally: photographs which bring out the conditions of work in prospective jobs and often stimulate comments that reveal the candidate's attitude; and trade test pieces which help to put the candidate at ease since he can talk about familiar, concrete things, and which bring out knowledge of methods of production, of standards of workmanship and so on.

Follow-up research on the Admiralty questionnaire, together with suggestions put forward by Wilson and Misselbrook, enable us to list a number of easily recognisable "contra-indications." When several of these bad points are present in the same candidate, they provide a poor prognosis for success at any responsible Service employment. Validatory investigations have yielded such uniform results in different jobs that we are entitled to regard "irresponsibility" as a general trait. It is the exception rather than the rule for a man who is a failure in one kind of work to find another kind to which he settles down and manages really successfully. The list includes:

(1) Unsteady, unprogressive or retrogressive work record; periods of unemployment considered in relation to district, also to ability. Thus, long-term employment at a low-grade job would be commendable in a very dull, but not in an intelligent, candidate.

(2) Inability to give an intelligible account of his own job.

(3) Post-war ambition below the level expected from pre-war occupation and test scores.

(4) Failure to reach the usual form for his age and intelligence at school.
(5) No further education in connection with occupation, or otherwise, since 14.
(6) Dislike at school for mathematical and science subjects. Handwork, athletics, geography or "none" listed as the school subjects liked best.
(7) Interests limited to the purely social and display type, e.g. jazz playing, the purely athletic, the entirely solitary, or too restless and migratory.
(8) No positive indications of leadership and responsibility.
(9) Evidence of marked pre-occupation with own health.

Overlapping with this list is the series of questions, drawn up by naval neuropsychiatrists, for application by Wrens at recruiting centres, in order to bring to light potential cases of psychiatric breakdown. The questions were not asked in a stereotyped fashion but were introduced, often indirectly, as topics for discussion.

(1) Lack of any sustained interests or hobbies, and avoidance of social or of physically dangerous interests.
(2) "For what illnesses have you been off work during the last five years?"
(3) "Has your health been good? Have you had to go to your doctor a certain amount (or often)?" The aim is to find how far minor ailments have been allowed to interfere with normal activities.
(4) Digestion, dieting. "Do you have to be careful in what you eat?"
(5) "How well do you sleep?"
(6) "What about your powers of endurance? Do you easily get tired?"
(7) "Would you regard yourself as sensitive or highly strung? Have you ever had any trouble with your nerves?"
(8) Incoherence of employment record.

A recruit whose illnesses interfere disproportionately with his career, or who drifts from one job to another for no good reason, or who shows others of these symptoms, was marked "P" by the Wren on his questionnaire—that is positive signs of instability.

Interviews by Army P.S.O.s had to keep more closely to the qualification form, since this had not been previously checked by recruiting assistants as had the naval questionnaire, also because it was desirable to restrain their desire to employ hunches in preference to factual evidence. They were, however, instructed to sum up their conclusions in the form of rough ratings for the following qualities:

(1) C.T. (combatant temperament). A low rating was given if the P.S.O. felt "apprehension about going into action" with the recruit, if his mode of life was unusually sheltered or restricted, or if he showed undue avoidance of vigorous and adventurous pursuits, and fear of serving abroad. (Such cases were referred to the psychiatrist.) A high rating was given to the outstandingly adventurous and aggressive.
(2) E.R. (employment record). A high rating was given if employment (considered in the light of aptitudes and opportunities) had been progressive, coherent and continuous; a low rating if the opposite.
(3) L. (leadership). This was based on good, average or poor experience of supervising and directing others in his work or spare-time activities.
(4) O.R. (officer recommendation) and/or potential N.C.O. were also recorded.

Little information is available as to the employment interviews conducted by psychiatrists either with the low-grades, Army officer candidates, or neurotics. However, Gillespie issued a list of factors to be considered by R.A.F. medical and psychiatric interviewers in assessing neurotic predisposition. These overlap to some extent with the points into which Service psychologists enquired:

(1) Symptoms such as stammering, persistent enuresis, insomnia.
(2) Neurotic fears—of the dark, loneliness, closed spaces, etc.
(3) Unsatisfactory record at school and work, the latter often shown by frequent changes.
(4) Comparative lack of interest in games involving bodily risks —boxing, diving, climbing.
(5) Inadequate reasons for joining R.A.F. or air crew.
(6) Response to sight of blood, e.g. fainting.
(7) Visceral responses before some ordeal such as an examination, important match or crucial interview.
(8) Family background, history of mental and nervous illness, "broken home."
(9) Reactions to fighting at school.

(10) Food, alcohol and tobacco habits excessively temperate or intemperate, or hypochondriacal.

(11) Temperament: Sociable or unsociable; obsessional traits; hysterical (dramatic, immature, suggestible, dodging); psychopathic (lack of persistence, extreme irritability, generally poor personality); depression; anxious; narcissistic (conceited, flamboyant, self-centred).

(12) Persistent air-sickness, headaches, etc., during training.

Reliability and Validity of Interview Judgments

Investigations of the employment interview described by Hollingworth (1929), Hartog and Rhodes (1935) and others suggest that there is so much divergence between the views of different interviewers of the same candidates, that the technique is practically worthless except for such restricted purposes as a business man choosing his own secretary. Several experiments, however, have shown much the same reliability for interview judgments as for ratings by friends and acquaintances, namely, correlations of between ·5 and ·6. Thus, in Webb's (1915) investigation, when schoolboys were interviewed only for a minute each by two interviewers separately, the agreement between their ratings on several traits was ·51. In Magson's (1926) research, boards of five interviewers, including business and professional men, assessed the intelligence of student teachers from interview with an average inter-correlation of ·52. Here one interviewer did all the questioning, the others listened. Additional judges who only read a transcript of the conversation correlated ·38 with the interviewers. Fearing (1942) studied ratings of police officer candidates by boards of four judges and found correlations of ·23 to ·48 on ten separate traits. But for the total grading based on all the traits the mean agreement was ·59. Several investigations of public opinion surveys show high, but far from perfect, agreement when a second interviewer asks the same questions shortly after the first one (cf. Cantril, 1944; McNemar, 1946). On straightforward, concrete questions such as ownership of a motor-car or telephone there was approximately 90 per cent. agreement, and on ratings of age 71 per cent. of identical judgments, corresponding to a correlation of ·91. But with ratings of more complex qualities such as economic status, only 54 per cent. of judgments were identical—a correlation of ·63. Perhaps the most striking study was that of Newman, Bobbitt and Cameron (1946), who interviewed 536 U.S. coastguard officer candidates and tried to assess the very complex quality of ability to pass training and to make a good officer in action. One psychiatrist's judgments correlated ·81 and ·86 with independent interview judgments by two psychologists. This suggests that psychologically-trained interviewers who have reached a clear and agreed conception of what they are looking for can achieve very satisfactory reliability.

Investigations of validity are not very favourable. Webb obtained a correlation of ·63 between his interviewers' judgments and judgments by teachers who knew the boys, but Magson only found a correlation of ·18 between interviewers and fellow students, and ·12 between interviewers and results of intelligence tests. Possibly the conception of intelligence among adults is particularly equivocal. Clark (1926) showed that two interviewers could estimate the college grades of students with validities of ·66 and ·73. But even these predictions are little if any better than might be obtained from a battery of aptitude and achievement tests, and from previous grades. The data reported by Bobbitt and Newman (1944) on 1,900 cases yield tetrachoric correlations of ·49 between combined interview judgments and passing the officer cadet training course. Test scores alone (which were known to the interviewers) had a predictive value of ·47. The combination of interview judgments + tests gave the best validity, namely, ·56. Anderson (1929) presents a number of tables showing the relevance of qualities assessed in a psychiatric interview to success at salesmanship and other commercial appointments, and claims that routine selection methods which omitted this interview were less diagnostic. Sarbin (1944), on the other hand, quotes studies by Wittman and himself in which predictions derived from batteries of tests or other factual items were superior to clinical judgments based on interview. Thus, in one research on the prognosis of schizophrenic patients, an appropriate combination of items from a rating scale on present symptoms gave better correlations with subsequent outcome than did predictions by psychiatrists. Similarly, in a study of college students, two tests (whose multiple regression equation had been established in a previous group) correlated more highly than interview judgments with academic achievement. This would indicate that suggestion, halo and bias, fallacious inferences from external signs and the like, play so large a part even in professional interviewing that the inclusion of an interview may distort rather than improve vocational procedures. Conrad (1947) reaches the same conclusion.

The controversy between the "clinical" and "psychometric" approaches to the study of personality is a perennial one. Allport (1942) contrasts the "nomothetic" and "idiographic" viewpoints in psychology. He claims that the objective and analytic methods of the former disrupt the integrated structure of the personality, and that the more intuitive idiographic study of the individual person as an organised whole makes for better "understanding, prediction and control." Hence the latter approach is no less scientific than nomothetism. Vernon (1935) describes experiments by von Bracken, Cantril and himself showing "that personality can be more accurately and consistently judged as a structured whole than can the separate traits or components into which it may be analysed." Similarly Polansky's (1941) investigation proves that better understanding of people and prediction of their actions may be derived from "structural" case studies than from lists of test scores. These results do not, however, affect the present issue—namely, that clinical judgments of the personality as a whole, based on interview, may be so unreliable or so distorted that better predictions of suitability for a particular vocation may be obtained from properly weighted combinations of psychometric data. Yet another aspect of the same controversy occurs in public opinion research. It seems very plausible that the free interview and observation techniques used by Mass Observation provide a better understanding of people's attitudes than do answers to a stereotyped list of questions, but it has yet to be proved that subjective factors do not play an unduly large part in Mass Observation's collection and interpretation of its evidence (cf. Durant and Harrisson, 1942).

As long ago as 1925 Viteles argued that an adequate picture of occupational suitability cannot be obtained from test scores or objective items of past history, and that the clinical approach of the trained psychologist is needed to take account of the other non-measurable factors and to integrate them into an over-all view of the candidate. Freyd (1925), however, maintained that the relevance of any such additional factors should be systematically investigated and validated in the same way as that of test scores. Wallin (1941) and Sarbin (1944) again state that the interviewer is really doing the same thing as the psychometrist—that is, combining the indications of various items which he believes, on the basis of past experience of similar cases, to be predictive of success or failure, into a total judgment of suitability; but that he is doing it informally and subjectively, and, therefore, less effectively.

Clearly it is incumbent on psychologists who regard the interview as a valid instrument of selection and guidance to produce experimental evidence. Members of the N.I.I.P., whose views accord with those of Viteles, have indeed proved the validity of their vocational guidance which largely involves interviewing, but have not demonstrated that they could not predict as effectively without it. Yet another important query that we must try to answer is whether an interview by a trained psychologist or psychiatrist is superior to one by an ordinary industrial employer or an experienced officer in the Forces.

Reliability of Service Interviews

No data are available comparing two independent interviews by psychologists or P.S.O.s except those listed in the previous Chapter (Table VI), but there is evidence of considerable variability in standards of judgment. Thus, the overall proportion of naval recruits marked "P" (unstable) by Wren recruiting assistants, who interviewed some 80,000 men in 1943, was 2·69 per cent. But the range of 101 different Wrens (each of whom interviewed 300 to 1,200 men) was from 0 to 12·9 per cent., and the quartile range 0·5 per cent. to 3·5 per cent. Each Wren had a fairly steady rate of her own, the correlation between the "Ps" in the first and second halves of the period being ·79. It was also found that Wrens judged by their supervisors to be high on such traits as carefulness, steadiness, conscientious, reliable, experienced, and teachable, marked the largest numbers. The rate varied considerably with area—that is, all the Wrens working in a given area had a common policy, but there were significant individual differences over and above this.

Such results should not be regarded as a serious criticism of the Wrens, since, owing to the impracticability of arranging psychiatric interviews of "P" men, little stress was laid on this rating. In H.M.S. *Royal Arthur*, P.S.O.s were encouraged to refer all doubtful cases to the psychiatrist, and among nine of them, each of whom interviewed 500 to 1,200 recruits during 1944, the referral rates ranged from 2·8 per cent. to 9·7 per cent., the total rate for over

8,000 recruits being 6·3 per cent. The psychiatrist classified these referrals as follows:

5·7 per cent. were considered as suitable only for immediate discharge on psychiatric grounds.

17·8 per cent. were judged likely to give six months' service or less before breaking down or showing criminal tendencies.

49·5 per cent. were borderline, being likely to give at least a year's service; they were, however, definitely thought worth referring.

27·0 per cent. were found to possess no psychiatric abnormality.

These figures suggest fairly close agreement between the P.S.O.s' and psychiatrist's judgments of instability, though in the absence of a control group of men *not* referred, no exact correlation between them can be calculated. The numbers in the first two groups referred by different P.S.O.s also varied from 11 per cent. to 37 per cent., and it was noted that, on the whole, P.S.O.s who sent the biggest total numbers also picked the biggest proportions of these serious cases. In the Army, P.S.O.s referred 14 per cent. of recruits to psychiatrists in a representative period, but their variability was not checked. Large variations were found, however, in the proportions of officer recommendations made by different P.S.O.s.

Evidence of the reliability of psychiatric interviews in the R.A.F. is given by Hill and Williams (1947) in a study of 541 cases who were seen by more than one of thirty-seven psychiatrists at different centres, at intervals ranging from several days to a few months. Among the items noted by the psychiatrists were diagnosis or "reaction type" under one of six headings, and degree of predisposition to neurosis—severe, mild or none. Of the diagnoses 81 per cent. were in exact agreement, but the figure varied considerably in different reaction types. If the average agreement in four main types is taken—anxiety, depression, hysteria, and others—it drops to 68½ per cent., corresponding to a reliability coefficient of ·76*. The reliability of judgments of neurotic predisposition was somewhat inadequate, corresponding to a tetrachoric correlation of ·58. Nearly half the men assessed as "severe" by one psychiatrist were called "mild," or occasionally "nil" by the other. Further judgments, such as degree of flying or other stress, showed intermediate degrees of agreement.

---

* This was calculated by Burt's (1945) matching formula.

Validity of Recruiting Assistants' and P.S.O.s' Interviews

The ability of Wren recruiting assistants, using the questions listed above, to pick psychiatric cases was studied in an experiment where four Wrens interviewed 147 actual psychiatric in- or out-patients mixed with 175 controls (Curran and Roberts, 1945). On the basis of their enquiries into pre-Service history only they picked 58 per cent. of the cases and 16 per cent. of the controls as suspect. But it was later discovered that of the 16 per cent. "false positives," most of whom were seen by a psychiatrist, over half did actually possess some considerable degree of abnormality. The Wrens missed a rather high proportion of the affective psychotics and neurotics, but managed to pick up twenty of the twenty-four hysterical patients. Obviously this experiment could not prove that the Wrens would have been equally successful in diagnosing breakdown *before* it occurred. Another weakness was that the four Wrens, though not necessarily the best available, were known to be generally good at their work, and others might not have done so well. It should also be noted that their success was not any greater than that of the various personality inventories which have been tried out as screening tests (cf. Chapter XV).

Variations in success of interviewing were shown when 1,631 naval radio and electrical mechanics and wiremen, who had been selected by thirteen different P.S.O.s, were followed up. The failure rates during their training averaged 24 per cent., but ranged from 18 per cent. among men selected by the most successful P.S.O. to 37 per cent. for men selected by the least successful. This study, however, is incomplete, since we do not know how many potentially suitable men were missed. Conceivably the P.S.O. of whose selectees only 18 per cent. failed was more "choosy" than the rest, and had she accepted larger proportions of her interviewees her failure rate might have risen. In an analogous investigation in the Army, nearly 3,000 W.O.S.B. candidates were "followed back" to the P.S.O.s (or other officers) who had originally recommended them. Although the numbers interviewed by any one P.S.O. were rather small, it was possible to show that there were significant variations among P.S.O.s both in the proportions of suitable recruits whom they recommended, and in the failure rates of their recommends at the W.O.S.B.s; also that the more "choosy" P.S.O.s did not necessarily pick better candidates.

In seven investigations it was possible to correlate both test results and P.S.O. judgments of suitability, based on interview, questionnaire and tests, with subsequent training course results or proficiency grades, and in only one of these did the judgments show better validity than some of the tests. Thus, among 730 Army personnel trained as drivers the correlation between P.S.O.s' assessments and pass or fail was ·234, while the validities of the Bennett test (S.P.2) and age (youthfulness) alone were ·294 and ·288. Gradings by a P.S.O. with engineering experience were available for 1,014 boys starting training as Army tradesmen; compared with practical marks over the next 1–3 years their validity was ·272, those of the Bennett and Squares (S.P.4) tests being ·294 and ·288. However, the gradings were not intended to predict general suitability for trade training so much as to advise the training schools on the best type of course for each boy, hence this comparison may not be entirely legitimate.

The final course results of 411 naval radar plot operators were collected (cf. p. 244), and the correlation of these with P.S.O.s' assessments was ·374. The correlations for T2, a scale reading test and a graph reading test were ·416, ·430 and ·400 respectively. If T2 is held constant, the partial correlations with the criterion, i.e. the extent to which the P.S.O.s or the other tests can improve on T2 are ·137 for the P.S.O., ·172 and ·193 for scale and graph reading. It should be noted that in this instance the P.S.O.s had rejected a (not very large) proportion of recruits as unsuitable for training; hence those that did go forward were more highly selected on P.S.O. judgments than on T2, also more selected on T2 than on the other two tests. Probably if allowance could be made for this selectivity, T2, P.S.O. and scale reading would all show almost the same validities.

Table XXXIX (p. 246) lists the correlations of P.S.O. judgments of 601 radio mechanics and 342 electrical mechanics with success or failure on course, together with the validities of several S.P. tests among some 300 and 200 of these men. Again the P.S.O.s tend to be surpassed by T2 and by information tests which are directly relevant to the jobs. But here, too, selectivity imposes a greater handicap on the P.S.O.s than on the tests. Similar investigations in the U.S. Navy yielding equally unfavourable results are described by Stuit (1947).

The most successful judgments were made by one highly experienced P.S.O. She graded 460 men who had been sent on course as safety equipment ratings, without seeing either the men or their test scores, but only from their questionnaires. It should also be noted that the relevance of particular questionnaire items had not been studied in any previous follow-up investigation. Nevertheless, the correlation between her judgments and subsequent course marks was ·554, whereas T2 alone gave a correlation of ·393 and no other combination of the standard tests would have improved appreciably on this. Her correlation with T2 was ·312, hence it was clear that she was not merely assessing general ability from previous education and occupation, but was effectively judging the character traits and interests relevant to the job.

Validity of Psychologists' Interview Judgments

The superiority of psychologists' interview judgments both to judgments by a board of naval officers and to predictions based on intelligence tests was indicated in a study of 503 R.N.V.R. cadets. The great majority of these were seen by Wilson, who used the methods outlined above, and the remainder by other Admiralty psychologists. T2 scores were available, but not the reports, specially prepared for the Admiralty Board on the candidates' previous record in the Navy. The boards, on the other hand, had as well as these reports, the psychologist's reports and T2s. Only 228 candidates were allowed to proceed for training by the boards and of these 61·4 per cent. passed. Of the 176 favourably graded by the psychologist 71 per cent., and of the fifty-two unfavourably graded 29 per cent., passed corresponding to a tetrachoric correlation of ·57. As none of the candidates unfavourably regarded by the board were sent for training, its predictions cannot be expressed as a correlation, and the two percentages 61·4 and 71 are not comparable. But if all candidates had gone forward, the superiority of the psychologist would almost certainly have been more marked*. Predictions within this group of 228 on the basis of T2 alone gave a validity coefficient of only ·24.

In a small W.O.S.B. investigation of officers who suffered psychiatric breakdown, psychiatrists, psychologists, tests, and non-technical board members were compared. Eighty-nine cases were traced who broke down within 1½ years after passing their boards, and the judgments made at the time of boarding were looked up. Neither their final board gradings nor the president's, or M.T.O.s' opinions of them were any poorer than of normal acceptances. The psychiatrists had recommended the acceptance of 71 per cent. of them, but had given a significantly larger proportion of adverse reports than to normal candidates. The psychologists (sergeants) had not seen the cases, but had studied their "pointers" material and had actually recommended the rejection of a larger proportion than the psychiatrists, namely, 48 per cent. Since, however, the psychologists tended to give more adverse reports on normal candidates also, the validity of their predictions was probably much the same. Intelligence test results were also sub-normal, 14 per cent. instead of the normal 6 per cent. having an O.I.R. of 4 or below. No one would expect such tests to predict breakdown, but in this investigation numbers are insufficient and the data too incomplete for it to be proved that psychiatrists are better than psychologists, or vice versa, or that either are better than tests. All three, however, are significantly better than "lay" interviewers or observers.

---

* Since the agreement between the board's and psychologist's judgments corresponds to a correlation of ·71, it follows that the psychologist's predictions were validated only in a highly selected group. Thus in a less selected group of candidates his validity coefficient would certainly rise. The low validity of T2 is also largely due to selectivity. Among an earlier, unselected, group of 415 candidates its validity was ·70.

Validity of Psychiatric Judgments

In 1941–2 several experiments were carried out by T. F. Rodger, Wittkower and other psychiatrists into the assessment of officers who were attending the Company Commanders' School near Edinburgh. The commanding officer and his staff were able to provide exceptionally thorough descriptions of the personalities and future promise of these officers at the end of their training. A psychiatrist interviewed each man for an average time of just under an hour, and wrote an independent diagnosis. On comparing the psychiatric and the school's opinions close agreement was claimed in eighty-five out of 100 cases. In a similar study of 223 officers certain intelligence tests were also given—Group Test 33, Matrices and/or Army S.P. Test 15. On matching the two sets of reports close agreement was claimed in 52 per cent., substantial agreement in 36 per cent., some discrepancy in 6 per cent. and divergence in 6 per cent. This matching, however, was subjective and it is impossible to say what degree of objective correlation it represents. It was shown that there were no significant differences in the agreement rate for the three psychiatrists who did most of the interviewing.

Elsewhere Bowlby carried out a similar, but better controlled, investigation of thirty-six O.C.T.U. cadets, to whom he gave the F.H.R. test and Matrices and a twenty-minute interview. He predicted their actual O.C.T.U. passing-out grades correctly in twenty-eight cases, and was seriously incorrect in only two cases. At that time, however, the cadets were very heterogeneous, having been selected by old-type interview boards, and it was found that the intelligence tests alone would have caught ten of the thirteen men who failed training. Thus, both the psychiatric interview and the tests achieved validity coefficients of around ·80.

In 1943–4 psychiatrists took part in the selection of Army parachutists and, in order to gain greater familiarity with the job, themselves underwent parachute training. Brief interviews and a shortened version of the W.O.S.B. "pointers" procedure were used. The trainees were graded 1 to 5 for suitability. On following up 1,492 cases, of whom 20 per cent. failed their first stage of training, it was found that 3 per cent., 7 per cent., 10 per cent., 23 per cent. and 46 per cent. of men in the psychiatrist's five grades failed. This corresponds to a validity coefficient of ·58. But on following up a year later the agreement between the grades and either wastage (on medical, disciplinary or training grounds), or promotion to N.C.O., was less close, corresponding to correlations of between ·2 and ·3 (cf. Expert Committee, 1947).

A small experiment in the R.A.F. is reported by Rollin (1944). Twenty-five W.A.A.F.s who failed trade training for reasons other than dullness, and 100 normal controls were interviewed, and use was made of Gillespie's scheme (p. 151). The most diagnostic symptoms appeared to be:

	Failures per cent.	*Controls per cent.*
History of instability in family	44	6
Nervous symptoms before enlistment (morbid fears, depression bouts, etc.)	56	6
Previous nervous breakdown	28	3
Invalidism suggestive of neurosis	24	4
Personality difficulties (over-aggressive, passive or inadequate)	28	4
Undue timidity, shyness, over-dependence or worrying ..	24	9
Unsettled home (e.g. due to unemployment)	16	3
Occupational instability	8	0

This analysis is, of course, *ex postfacto*, and its validity could only be proven if it was tried out on fresh cases before their training started. The relation of neurotic disposition to operational efficiency was shown very clearly by Reid. Either the medical officer or Reid himself interviewed 200 airmen on arrival at a R.A.F. station, and they were followed up through a tour of operations. Two or more neurotic symptoms were found to occur in 20·4 per cent. of those who had accidents or became casualties, and in 69 per cent. of those who developed psychiatric breakdown, but only in 5·4 per cent. of those who successfully completed their tour. This corresponds to a tetrachoric validity coefficient of + ·55. Finally, a large-scale trial of the predictive value of psychiatric interviews was carried out in the R.A.F. towards the end of the war. Of the 500 men seen by two psychiatrists, only a few could be followed to the operational stage. The main finding was the extent to which the validity of the judgments depended on the personal flair of the psychiatrist (Williams, 1947).

Conclusions

While none of the experiments summarised above provides a definite answer to our problem—whether the "clinical" approach of the interview yields better predictions than does the psychometric approach based on an appropriately weighted battery of tests and factual biographical items—yet they do show that interviews by trained psychologists and psychiatrists possess considerable validity. Moreover, when adequate consideration has been given to the qualities to be assessed and the methods used, and when a certain amount of objective information is available as a starting point, the consistency or reliability of judgments by such trained personnel is reasonably satisfactory. At the same time the correctness of the predictions never approaches anywhere near 100 per cent., except perhaps, when the candidates are abnormally heterogeneous. It is noteworthy also that even professional interviewers probably do not allow sufficient weight to objective factors. Although they consider that they have taken test scores and the like into account, better predictions are often obtained from a combination of their final judgments with test scores than from either alone.

Psychological and psychiatric interviewers are almost certainly superior to experienced but psychologically untrained persons such as employers and officers in the Forces. It is doubtful whether interviews by P.S.O.s (even when carefully selected and trained) constitute a valid vocational technique. Probably some P.S.O.s are quite as good as psychologists and psychiatrists, but there are such large variations that it would seem wise to keep their diagnostic functions to a minimum and to rely to a greater extent on psychometric predictions whenever possible. *None of the evidence that we have collected suggests that their subjective summing up of past history and their personal judgments are capable of covering all the relevant factors in vocational classification omitted by tests.*

As shown in the previous Chapter, the interview is essential for checking the correctness of questionnaire entries, for giving information, for matching supply and demand, for improving morale*, and often because it is a vested interest. Doubtless it has valuable applications also in treatment, counselling, research, etc., and in providing hypotheses as to the occupational relevance of various symptoms—hypotheses which should be subjected to empirical verification. But we are forced to conclude that its value as a diagnostic technique is, as Hunt, Wittson and Harris (1940) put it, a matter of economics. Although thorough psychometric procedures might usually be more accurate they would involve such extensive and complicated research, and be so costly to apply, that they may be impracticable, in addition to being unacceptable both to candidates and to employers. The interview has the tremendous advantages of inclusiveness, speed and flexibility. It takes into consideration a much wider range of factors than can readily be tested (though admittedly many of these are irrelevant and most of them liable to misinterpretation). It requires no apparatus, not even printed blanks. In the U.S. Navy, according to Hunt, interviews averaging only three minutes in length, served for screening psychiatrically suspect recruits. It was found that a single psychiatrist could deal with 100 cases a day, though many could not keep up this pace for long. Even with ten- to twenty-minute sessions, an interviewer is likely to cover fifty men in no more time than a tester would take to give, score and weight the appropriate score combinations on, say, a two-hour battery, which would have required months of technicians' time to prepare. True, these interviews might make more mistakes than the objective procedures, but the sufferers (whether employers or employees) tend to resent errors of human judgment much less than they do errors made by impersonal instruments.

---

* For example, Closson and Hildreth (1944) have shown that recruits who receive a five-minute psychiatric interview, designed to discover their "weak spots" and to give advice accordingly, make better adjustments in the Forces than comparable men who have merely been selected by impersonal test methods.

The interview again can readily adapt itself to the exceptional case, e.g. the illiterate whose test results or questionnaire responses are quite misleading. It can be applied in employments which involve small numbers, where adequate validatory evidence for psychometric procedures is unobtainable, and in such fields as managerial, officer or air crew selection, where it is particularly difficult to devise tests of character or temperamental qualities. Finally, when the requirements of any employment alter, the interview can be adjusted at once, whereas the psychometric approach involves a complete re-investigation of the validity of all the possibly relevant test scores and biographical items, a fresh statistical analysis and preparation of a fresh weighting system. As we have seen, such re-investigation cannot properly be carried out on a selected group; it is hardly feasible to work out minor adjustments in a battery while that battery is in actual use for selection. For all these reasons, the interview rather than tests must remain the prime instrument of vocational classification, and vocational psychologists are justified in spending as much of their time on attempts to improve the technique of interviewing as on devising more valid tests.

CHAPTER X

PRINCIPLES OF PSYCHOLOGICAL TESTING

Abstract.—A psychological test, by presenting a standardised task or situation, elicits a sample of the testee's behaviour which can be objectively scored and compared with norms of performance, and which has been proved to be predictive of future occupational or other behaviour. To regard it as measuring particular traits or abilities (the "naming fallacy") is often misleading. Its content is best determined by a combination of systematic introspection, empirical validation, and factor analysis. The types of tests found most useful in vocational classification are group paper-and-pencil tests for measuring the factors underlying a large number of jobs, information or trade knowledge tests, and supplementary "miniature situation" tests, each closely analogous to some important job.

Given a standard battery of group tests, the planning and construction of new tests is guided not so much by job analysis of the aptitudes involved as by gauging how far this battery needs to be supplemented, and by considerations of practicability, i.e. the availability of materials and time for devising, giving and scoring, etc. The main steps in such construction are outlined. Adult tests are generally calibrated (provided with norms) in terms of percentile levels, rather than in mental ages, I.Q.s, or "critical scores." There are great difficulties in indicating the relevance of scores, especially on a battery of tests, to particular jobs. So-called "wastage coefficients" and minimum standards are unsatisfactory. Tables or histograms which show the optimum score levels on the most valid tests, and the probability of job success at each level, are advocated.

The fundamental method of science is observation of the way things (living or non-living) react in specified situations, and this is the method of psychological testing. The tester does not deduce people's traits or abilities from bumps on their skulls or from their physiognomy, nor does he rely on hunches or subjective impressions, but records their reactions to a standard task. At the same time a test is in no sense a miraculous instrument for revealing otherwise unsuspected qualities; it is simply a refinement of the process of observing behaviour which we use every day in making judgments about our acquaintances. As pointed out by Vernon (1940) this refinement involves:

(1) Standardising the task or situation, so that all testees react to the same task.

(2) Objective and unbiased recording of the testees' performance, usually in quantitative terms.

(3) Availability of norms or standards by comparison with which the goodness or poorness of performance can be assessed.

(4) Demonstration that the test is both reliable—i.e. a precise measure, and valid—i.e. that it actually measures or predicts what it claims to do.

Certain other points should be noted:

(5) This view of tests includes intelligence or educational tests, where the recorded behaviour consists of answers to intellectual problems, as well as practical vocational tests, also medical tests of eyesight, hearing, etc., which may have a bearing on vocational suitability. It does not cover so-called projection tests or other clinical techniques whose object is to provide the psychologist with insight into the testee's personality (cf. Thematic Apperception, p. 57), or to allow him to judge the testee's methods of work (cf. Cube Construction, p. 32). Such instruments are valuable in skilled hands, but are not strictly tests, unless objective scoring methods are applied.

(6) A test constitutes only a relatively brief sample, or set of samples, of behaviour. To make sure how well a man will do in a certain occupation we should need to let him try it in real life, but because this would take so long we employ the short cut method of tests. It naturally follows that no test, nor even an appropriately chosen battery of tests, will give perfect predictions of proficiency at the occupation. Both the sample behaviour elicited by the test and vocational proficiency itself depend on a multiplicity of factors, so that there can be only a certain statistical probability that testees obtaining such and such scores will succeed or fail. If the test has good validity the odds may be very high, but there

can never be complete certainty about any particular individual.

(7) So far as possible a test should be named in terms of the operations it actually involves, rather than in terms of traits or abilities which it is supposed to measure. It is scarcely possible to discuss or analyse vocational suitability without bringing in personal qualities like intelligence, attention, dexterity, persistence, etc.; but we should realise that these are hypothetical constructs. Such terms, which are intermediate between the observed test behaviour and the predicted future behaviour, are often unnecessary and misleading. The "naming fallacy" has been the bane of vocational psychology, witness the ubiquitous use of "manual dexterity," "visual discrimination," "attention to details," and the like. Factor analysis investigations indeed show that there is so little overlapping between different tests of dexterity that it is doubtful whether it should be credited with any general and independent existence. Such analysis does, however, suggest that certain abilities are sufficiently consistent and distinctive to constitute useful concepts for the psychologist's vocabulary, and these are usually denoted by letters such as *g*, *k*, or *v*, rather than by names of dubious significance such as intelligence, spatial or verbal ability.

There are, in fact, three methods of deciding what a psychological test measures, all of which are usually employed:

(i) The method of *a priori* or theoretical analysis. Inspection of a test merely gives us what is called its "face value," which has little or no scientific worth. The collation of systematic introspections about the test from several trained observers may be more useful, but even the views of psychologists as to the physical or mental processes involved in taking the test may be invalid when the prospective testees are children, or adults of sub-average intelligence. Chapters XII–XV give examples both of tests which failed to measure the abilities anticipated, and others which measured unsuspected abilities.

(ii) Follow-up evidence, i.e. the comparison of test results with success or failure in various jobs. Though this is the most scientific method, it too has its weaknesses—the difficulty of securing reliable criteria of proficiency, the problems introduced by dealing with selected groups, and so forth.

(iii) Factor analysis, which indicates the main types of ability involved in a test by analysing its correlations with other, partially overlapping, tests. When representative groups of adults are tested, ranging from the very highly intelligent and well-educated to mentally-defective level, it is found that a general or *g* factor enters to a large extent into practically all tests. Next the majority of tests fall into one or other of two contrasted types—the verbal, numerical and educational type (*v : ed* factor) and the practical, mechanical and spatial type (*k : m* factor). In more detailed investigations, these types break down into numerous sub-types or group factors. The former includes distinguishable, though usually overlapping, verbal, numerical, clerical and secondary school (as contrasted with primary school) abilities. The latter group includes mechanical, spatial, informational, physical, manual, and other abilities. The extent to which a test depends on any of these abilities, i.e. its factor loadings or saturations, may be expressed by correlation coefficients*. Such figures should, however, be accepted only with the greatest caution, since they are largely dependent on the heterogeneity or range of ability in the testees studied, and on the particular set of tests chosen for analysis. Further investigation might frequently reveal the presence of additional group factors, and, as Spearman showed, every test measures something specific to itself.

Varieties of Psychological Tests

Several classifications of tests have been proposed, and it will be useful to indicate which of the different types find most scope in vocational procedures such as those developed in the Forces.

Group v. *Individual*.—In large-scale vocational work, little individual testing is possible, and as big numbers are needed for trying out and validating any test (whether it is to be rarely or frequently applied), the tendency among vocational psychologists is to convert more and more of their instruments into group form. Even the interview, for example, is partly displaced by the biographical questionnaire which can be filled in by a group under supervision. Mechanical assembly tests employed in the Army and A.T.S. were adapted so that eight testees could be dealt with simultaneously. With sufficient ingenuity there is no reason why the majority of performance and apparatus tests should not become automatically self-scoring. Another reason for this tendency is that fully qualified psychologists cannot afford to spend much time on testing; tests, and their scoring, must, therefore, be simplified as much as possible, and the personal element in individual testing reduced, so that accurate testing can be done by less highly-trained individuals (cf. U.S. War Department, 1946).

---

* The squares of the correlations provide a better index of the make-up or content of tests. Thus in the case of Matrices with *g* and *k* : *m* loadings of ·79 and ·15, the variance attributable to these main types of ability is 62·41 per cent. and 2·25 per cent., leaving 35·34 per cent. unaccounted for. This amount represents group factors that have not been isolated, together with the test's unreliability and specificity. The actual factor loadings of many of the chief naval and army tests are listed by Vernon (1947b).

Paper-and-Pencil v. *Practical*.—The former includes, besides verbal intelligence, educational or trade knowledge tests, nonverbal *g* and spatial judgment tests based on abstract diagrams, and others—particularly in the mechanical field—based on pictures. Some tests are intermediate—for example, morse or code-learning tests—involve the practical discrimination of auditory patterns and written responses. Owing to the difficulties of constructing reliable apparatus or performance tests in war-time, and of maintaining sufficient sets in standard condition for application to thousands of recruits a week, there was a tendency to rely very largely on paper-and-pencil tests. This occurred, too, in America, where the difficulties were less acute. Commenting on the same tendency after the first World War, Hull (1928) surmised that a far greater variety of behaviour could be sampled by performance and apparatus tests. Admittedly there may be some danger of handicapping the "handyman," or the recruit with manipulative skills who, perhaps through no fault of his own, left school early and is inefficient in reading and writing. Whether we are favouring the "bookish" man in all jobs, both in the Forces, the Civil Service, the police, industry, etc., is a problem that needs serious investigation. But the danger should not be exaggerated, since "practicalness" is another rather dubious construct or hypothetical trait. Certainly a man may do better on a performance test of intelligence than on a written one, or on a manipulative assembly test than on a pictorial mechanical one. But there is so little evidence as yet of the superiority of such a man in engineering or other practical jobs that the introduction of practical tests may not be worthwhile from the economic standpoint, though, nevertheless, often justifiable on the grounds of "sales value," i.e. attractiveness both to testees and to employers.

Paper-and-pencil tests may be subdivided into "expendable" and "non-expendable." In the former the test blank itself is used, in the latter answers are written or (as in American machine-scored tests) marks are made in the appropriate positions on separate answer sheets, and the test blanks themselves can be used many times over. Non-expendable tests are often preferred because they save paper and are more easily scored, but they undoubtedly impose an additional handicap on duller testees. The creative or inventive response type of test item is also more suitable for such testees, and was largely employed in British tests because multiple-choice or selective-response tests and examinations are still so unfamiliar in this country. But it is difficult to construct inventive items which have only one right answer. Various alternatives are possible both in the information and mathematical tests described below. Even when all permissible answers are listed, there is still some danger of differences in scoring between different testers. American psychologists claim an additional advantage for selective-response tests, that the testee does not have to phrase his answer and can, therefore, express his knowledge or ability more directly. Such arguments have been critically discussed by one of the writers elsewhere (Vernon, 1940).

Aptitude v. *Attainment Tests*.—The contrast between inborn and acquired ability is less stressed in present-day psychology than in the past, since it is realised that every test involves both. A relative distinction is useful, as Traxler (1946) puts it, between tests based on tasks in which there has been little or no formal training, and others based on tasks similar to those that have actually been studied. Thus, vocabulary constitutes a legitimate test of intelligence because knowledge of words is chiefly picked up incidentally. Information tests such as S.P. 117, designed to show how much general mechanical and electrical information a man has acquired from hobbies, from doing odd jobs at home or handwork at school, or from reading (but which exclude items involving specific trade experience) appeared to provide better indications of interest in, and aptitude for acquiring further skill in, mechanical jobs than did any tests belonging to the more conventional aptitude category. Similarly in predicting university achievement, it is often found that previous attainments and tests of such generalised study skills as reading comprehension are of more value than intelligence tests (cf. Vernon, 1939a; Eysenck, 1947b).

Analytic, Analogous and Work-sample Tests.—This classification of vocational tests overlaps closely with the previous one. Analytic tests of the elementary sensory and muscular functions believed to be involved in a job, or analogous tests which parallel the main job operations, are the generally accepted tools of vocational selection, since they are supposed to be uninfluenced by the unequal experience or training candidates may have had. While numerous investigations favourable to such tests have been reported, particularly by German industrial psychologists, they do not bulk largely in most modern vocational schemes, for several reasons.

(1) There is clearly considerable truth in the claim of Gestalt psychologists that every industrial operation is a whole which is more than the sum of its elements. Proficiency depends less on the sensory-motor capacities covered by analytic tests than on the successful integration of such capacities. Analogous tests also often give poor predictions because the structure of the abilities which they test differs in some essential respect from the structure of the job itself.

(2) The analytic or analogous approach is extremely susceptible to the "naming fallacy."

(3) Vocational classification usually has to cover so great a variety of different jobs that it would be impossible to devise separate batteries of tests even for the main ones, or to put all the candidates through several batteries. The only practicable approach is to test the common factors or generalised abilities which enter into a large number of jobs by a limited battery of tests, and then, perhaps, to add one confirmatory test in each main job which can be applied to likely candidates for that job.

(4) This approach is justified by the findings of factor analysis. In a heterogeneous population such as recruits or school-leavers, factors such as g, $v : ed$ and $k : m$, when supplemented by biographical data, cover a major proportion of the ground, and the addition of specialised tests for particular

jobs adds relatively little to the success of predictions. There may be more justification for applying batteries of specialised tests in vocational selection where the candidates form a more homogeneous group (cf. Chapter XII).

"Work-sample" tests appear to be differently interpreted by different industrial psychologists. Thus Viteles (1932) implies that work-sample testing involves giving the candidates some preliminary training on the job itself and deducing their eventual proficiency from their initial results or rapidity of learning. Others use the term of any trade test such as the dictation of a standard passage at standard speed to test stenographic ability. Several of these were widely used in the Forces for evaluating the trade experience of recruits.

We may conclude that the most appropriate tests for vocational classification are of three kinds:

(i) Group tests for the main factors underlying proficiency;
(ii) Work-sample tests and tests of trade knowledge;
(iii) Supplementary tests intermediate between the analogous; and the work-sample type, which attempt to reproduce, as it were in miniature, certain essential features of the job not covered by (i). Of those described below Agility (Test 16), Morse Aptitude (Test 10), Asdic gramophone records and the N.I.I.P. clerical test, come under this heading. The fact that these were to some extent influenced by training or experience did not greatly detract from their value, both because recruits with relevant experience were found to be so much more readily trainable (cf. Chapter VIII) and because, as pointed out above, attainment generally gives useful predictions of aptitude. Moreover, there was the safeguard that the P.S.O. could interpret the test scores in the light of opportunities. A high score on, say, a clerical test would be more significant in a recruit whose civilian job was non-clerical.

Planning and Construction of Tests

The traditional approach of the vocational psychologist is to start with a detailed job analysis of the methods of good and poor workers at the job, to construct a large battery of analytic or analogous tests for measuring the abilities so revealed, and to validate this battery empirically (cf. Bingham and Freyd, 1926;

Hull, 1928). We would suggest that this applies only in vocational selection schemes whose scope is limited to one or two particularly important jobs, the applicants for which are fairly homogeneous in *g*, education and experience. Certainly the psychologist concerned with vocational classification should try to study workers on the job, and should consult supervisors, training staff, etc., regarding the nature of the work, but with a rather different aim in view. First he needs to know as much as possible about the physical and social conditions under which the job is performed, and, secondly, he should try to assess what qualities are *not* likely to be covered by his standard tests and by biographical information. In radar and asdic operating, for example, even when these have been given full weight, there are likely to be left over certain visuo-perceptual and auditory capacities which do demand special tests. Both jobs, too, appear to demand a lot of knob-twiddling and concentration of attention, but the experienced psychologist recognises that the effort to try to measure such qualities would probably be wasted. If, as in these instances, additional tests appear desirable (or if follow-up evidence shows that his predictions are insufficiently accurate), these tests are designed, not so much to measure hypothetical aptitudes as to duplicate as closely as possible the remaining important aspects of the work. Thus, in planning tests for asdic operators (cf. p. 26) it was mistaken policy to list sense of pitch, sense of timbre, auditory threshold, etc., and to apply tests like Seashore's which claim to measure these. In fact, such tests merely acted as rather inefficient tests of general intelligence. But by constructing gramophone records with asdic noises which resembled the discriminations needed in the job itself, which were also so simple that their dependence on *g* was low, a distinct contribution was made to the prediction of operating efficiency.

In the planning of tests for vocational classification, there are two factors which are on the whole more important than detailed job analysis, namely, the experience of the psychologist, and practicability. The psychologist needs to know what kinds of tests have worked well in the past in selecting for similar jobs, and should be familiar with the evidence from American and other relevant literature.

The economic aspects of testing have been recognised by many writers, though hardly sufficiently stressed. Hull admits that progress in testing is likely to depend as much on the reduction of testing costs and the development of mechanical methods of scoring and of weighting scores, as on the predictive value of the tests themselves. Viteles points out that it is not worth while embarking on a programme of testing unless a sufficiently large number of sufficiently co-operative workers can be ensured on whom to try out and validate the tests, and unless a satisfactory criterion of their proficiency is available. In the Forces, for example, it appears so difficult to obtain an adequate measure of officer ability (nobody knows, or no two authorities agree, as to just what constitutes a good or a poor officer) that it is hardly worth attempting to devise tests of officer aptitude. Other important considerations include:

(1) What kind of tests can be given and scored by the available staff in the available time, both at the experimental trial stage and in routine practice? Most of the Service tests listed below are very short (ten minutes or less) because of the restrictions on testing time and because of the danger of recruits becoming "fed up" if over-tested. Even the effects of practice which would arise if recruits took several hours of tests are shown to be serious, in the next Chapter. As much relevant information as possible should be obtained from questionnaires and interviews which do not *look like* tests, even if, in fact, the answers are standardised and validated in such a way as to constitute additional tests. Headmasters similarly tend to resent allotting more than an hour or two to the testing of school-leavers. Promising tests of interests and of neurotic tendencies could not be introduced in the Army because their scoring (in the absence of machine methods) was too lengthy.

(2) Will the paper or other materials be forthcoming in the future as well as at present? Can apparatus be maintained in proper working order? The paper shortage in the war considerably restricted the use of tests containing spatial or mechanical diagrams and pictures, and the diversion of almost all skilled mechanics into the Forces or into armaments production rendered it virtually impossible for psychologists to adopt any tests involving elaborate apparatus. Even a very simple test of visual acuity which was applied to all A.T.S. recruits during one period had to be abandoned because it was yielding utterly inconsistent

results in different places. A complex reaction time test for motor drivers (Chleusebairgue, 1939), "miniature situation" tests for layers and for radar operators, tests of ability to talk over the telephone, and tests of night vision (as distinct from dark adaptation) were others which were regarded as valuable but impracticable.

(3) Can the psychological and statistical staff cope with the work needed in devising, analysing and calibrating a new test? Is it likely to be sufficiently superior to older established tests to justify this expenditure of technicians' time?

(4) The range and general level of intelligence in the population for which the test is intended must also be borne in mind. Some tests which might be quite practicable at upper levels may need far-reaching modifications and experimentation to be got across to duller men. Without this they may merely measure *g* over again, or be highly susceptible to practice effects. A warning has already been given against separate answer sheets and selective-response items. Figures, diagrams and drawings need to be large and clear; thus, here, too, lack of paper may be a limiting factor.

The production either of a general or a supplementary test is a highly complex and technical business, and we can only touch on a few points here. When the purpose and nature of the test have been decided, suggestions for items should be collected from several psychologists or other sources, and they should be carefully scrutinised and revised to eliminate ambiguities, to ensure technical accuracy, and to cover the requisite range of difficulty. A large excess of spare items should be included, and if alternative forms are likely to be needed, it is preferable to devise and try them all out simultaneously. Suitable instructions, for the testees and for the testers, must be prepared and time limits decided, subject to modification in the light of preliminary trials. Usually at least two large-scale trials, on populations which number several hundred and which parallel the populations for whom the test is intended, are needed for the following purposes:

(1) To see if the layout and instructions are adequate, and whether there are gross defects in any items.

(2) To validate the test as a whole and to show that it adds appreciably to other existing tests, or is superior to other proposed tests.

(3) To establish the difficulty, also the consistency and/or validity of each item.
(4) To validate the final form or forms and determine its predictive value in one or more jobs.
(5) To find its reliability.
(6) To analyse the main factors it involves by comparison with other tests.
(7) To calibrate, or determine its norms.

Calibration

Neither the obtained "raw" score, nor the percentage score on a test or examination provides a precise indication of the goodness or badness of performance. The results of intelligence or educational tests among children are, therefore, often expressed as mental or educational ages. Though it may be useful to know that, say, the arithmetic of a poorly educated adult is only equivalent to the level achieved by average ten-year-old children, age units are scarcely applicable to superior or even to average adults. Abilities increase or decline after fourteen years in so varied a manner (cf. Chapter XI) that no fixed standards can be determined. Intelligence quotients or I.Q.s are sometimes used with adult intelligence tests, for example, the Wechsler-Bellevue scale, but are disliked by vocational psychologists not merely because their derivation differs from that among children*, but also because comparable units for other tests are not available. For example, it might be more confusing then helpful to turn mechanical comprehension or morse aptitude test scores into mechanical quotients and morse quotients, and so on.

In general, therefore, adult test results are interpreted by percentiles. For example, a score of 46 on the (twenty-minute) Matrices falls at the 90th percentile for representative adults, since 90 per cent. score 46 or below. Similarly a score of 36 is the median or 50th percentile. These percentile levels can be based on any convenient group. Thus, in the Navy different percentile norms were available for ordinary seamen and for R.N.V.R. officer cadets. As, however, the percentile scale is unnecessarily detailed for most purposes, both the Navy and Army divided the ranges of scores on most of their tests into selection grades or groups (S.G.s).*

---

* Instead of being based on the ratio of mental to chronological age, they are usually obtained by converting test scores to an arbitrary scale with a mean of 100 and a standard deviation of 15 or 16½, or 20, etc.

Group A or S.G.I. represents the top 10 per cent. of scores, i.e. the outstanding.

Group B or II represents the next 20 per cent. of scores, i.e. above average.

Group C or III+ and III− represents the next 40 per cent. of scores, i.e., average.

Group D or IV represents the next 20 per cent. of scores, i.e. below average.

Group E or V represents the bottom 10 per cent. of scores, i.e. very low.

If the same grouping was applied to the Terman-Merrill revision of the Stanford-Binet test the I.Q.s falling into these various grades would be approximately: S.G.I 121 and upwards, S.G.II 109–120, III+ 100–108, III− 92–99, IV 80–91, V 79 and below. Such a system is easily got across to the layman, and has the great advantage that the S.G.s or the percentiles for all tests standardised on the same population are comparable. A 70th percentile score, for example, represents the same relative degree of superiority on Matrices or mechanical comprehension or morse aptitude.

Yet another form of calibration sometimes used in vocational work consists of minimum or critical test scores needed for entrance to certain jobs. This is an undesirable system for several reasons. First, it neglects variations in supply and demand. Obviously the critical score can be raised when there are plenty of candidates for a few vacancies, and must be lowered if the situation is reversed. Secondly, such rigidity is essentially unpsychological. We have shown that test scores constituted only one of the many factors influencing employment recommendations. P.S.O.s were encouraged to select men with lowish scores who possessed other outstanding qualifications, and to raise their standards if they did not. Thirdly, there are numerous technical difficulties in fixing satisfactory pass marks (cf. McClelland, 1942), particularly when several tests need to be taken into account.

---

* Other types of standard scores are often used in America but appear to offer no particular advantage. *Stanines* group scores into nine levels instead of our five SGs. A 7:24:38:24:7 per cent. split is often preferred to our 10:20:40:20:10, since it accords more closely with the normal frequency distribution, but it is less readily grasped by the non-psychologist.

Interpretation of Test Results

Although we have rejected the notion of critical scores, we cannot shirk the problem of providing some method of indicating the relevance of test performances to jobs. Not only the P.S.O. but also the interested layman wishes to know how accurately or inaccurately a battery of tests will select, and at what level of scores candidates are likely to succeed or fail. The only method which satisfies the statistician is that of correlation and regression coefficients, but, unfortunately, it is hardly intelligible to persons without statistical training, and is often scarcely practicable. Most of the proposed alternatives such as "wastage coefficients," "efficiency indices," etc., are very cumbersome and tend to be misleading. For example, a common procedure is to compare the distributions of test scores of men who turn out to be good and bad at the job, and to choose a pass mark or critical score which admits as many of the former and rejects as many of the latter as possible. This is reprehensible since such distributions (unless obtained from thousands of cases) always show irregularities, and one is tempted to choose a mark which, by taking advantage of irregularities, gives the most favourable differentiation. In fact, the smaller the groups compared, the more optimistic is the forecast of the test's validity likely to be. Moreover, the numbers or proportions thus found to be correctly selected or rejected depend to a large extent not merely on the goodness of the test, but also on the proportions of goods and bads, and on the proportion of candidates which P.S.O.s can afford to exclude. As pointed out in Chapter VII, the higher the selection ratio, the better will the selection test appear to be—provided that candidates wrongfully excluded by it are conveniently forgotten. Other common procedures are to list the percentages, or to draw graphs or histograms, showing the proportions of men at various test score levels likely to do well or badly at the job. These, too, suffer from the defect of being incomparable with jobs that have different failure rates. A slightly more satisfactory system was adopted by the U.S. War Department (1946), and is illustrated in Fig. 3. This shows that candidates scoring 140 on the clerical aptitude test had 84 chances in 100 of achieving *better than average* results at a training course for clerks, whereas candidates with scores of 60 only had 7 chances in 100 of doing as well. This method can at least be applied uniformly with all tests and all jobs, but it does not readily indicate how low the P.S.O. can go in selecting men unlikely to fail the course. Nor, when several tests are predictive of success, does it indicate how their findings are to be combined, or how much attention should be paid to each.

In the British Army, S.G.s were suggested for each main job on the tests believed, as a result of job analysis, to be most relevant, these being modified appropriately in the light of distributions of scores among men actually engaged on the work, and of later validation data. Thus, a minimum of S.G.3—might be set on the clerical test, but no minimum on the mechanical comprehension test, for clerks. Though this system is generally applicable and adequately flexible, it was unsatisfactory for several reasons.

(1) No information was provided as to the likelihood of success or failure above or below the minima.

(2) The imposition of different minima on different tests is not, in fact, equivalent to weighting the tests in accordance with the obtained regression equation.

(3) Some P.S.O.s found it too complicated and preferred to rely on their own hunches, but others followed it too literally, with disconcerting results. For example, if no standard was set on one test, say arithmetic, most of the weak arithmeticians would be herded into this job.

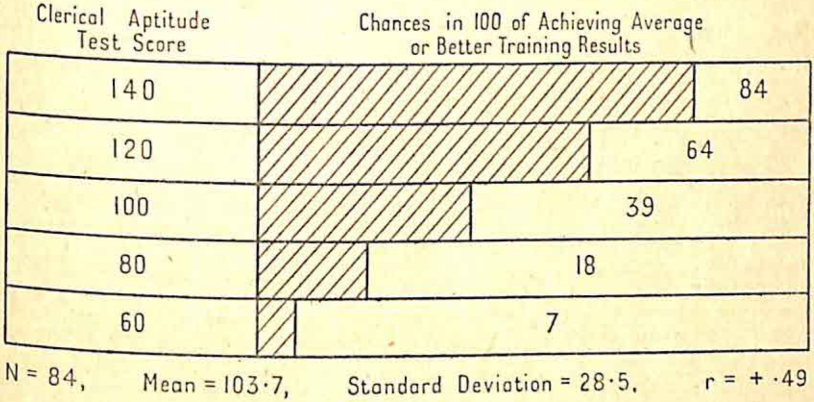


FIG. 3. PREDICTIVE VALUE OF A CLERICAL APTITUDE TEST.

It is the writer's conviction that any such mechanical system of test interpretation would soon be liable to develop similar flaws. A scheme which was developed in the Navy gives the distributions among passes and among failures (or men regarded by instructors as very poor), on two or three of the most relevant tests only.

Usually one test is the measure of all-round ability, T2 and the others are supplementary ones to which special attention should be paid*. As an example, Table XII lists the information for coders.

TABLE XII.—SELECTION TEST DATA FOR CODERS
Period: July-Sept. 1943, $N = 528$

S.G.	*T2*		*Test 1 Abstraction*		*Test 70 Dictation*	
	Pass	*Fail*	*Pass*	*Fail*	*Pass*	*Fail*
A	31	½	29½	½	21½	1
B	41	2½	49	3	40½	3
C	25	7	19	7	37	8
D	2½	2	2½	2	1	½
E	½	½	—	—	—	—
90 percentile ..	125		16		35	
Median	100		13		33	
10 percentile ..	72		10		30	
Correlation with practical training	·543		·584		·370	

The figures could equally well be presented in the form of a histogram. The pass columns give the percentage S.G. distributions (to the nearest half) among trainees found satisfactory. The fail columns give the proportional numbers. Thus, the actual failure rate was 11·5 per cent., corresponding to 12½ fails for every 100 passes. Percentile levels among passes are also shown. Thus, the P.S.O. can readily see that As and Bs on the three tests have excellent chances of passing, that Cs are doubtful bets, and that only Ds with outstanding additional qualifications are likely to be suitable. At the same time as indicating desirable standards for selectees, this scheme provides evidence of the value of the main tests and involves no statistical manipulation of the empirical data which would be liable to misinterpretation. It might well be adapted to civilian educational and vocational classification procedures. For example, similar tables might show the adequacy of a child's scores on general intelligence, educational and mechanical tests in relation to academic or to technical secondary schooling.

---

* In most naval jobs it was found that a very close approximation to the multiple regression equation could be obtained by T2 alone, or by T2 plus one or two supplementary tests equally weighted.

CHAPTER XI

EFFECTS OF PRACTICE, AGE AND OTHER FACTORS ON TEST SCORES

Abstract.—Investigations of civilians (adults and children) and of recruits show an average rise, on taking a psychological test a second time, equivalent to about 5 I.Q. points, apparently regardless of the proximity of the re-testing. Further re-tests produce progressively smaller increases, hence the provision of practice sheets or of one or two preliminary tests, helps to even up testees who possess different degrees of previous experience. The effect is smaller in straightforward tests with ample time limits, than in choice-response tests or tests with complicated instructions, with which most testees in this country are unfamiliar. Though the practice effect of one test on another is about twice as great when they are identical as when they are only partly similar in form or content, yet any kind of schooling and the taking of examinations may produce slight improvements among adults.

Recruits did not constitute sufficiently representative samples of the general population to provide definitive results on changes in test scores with age. However, the more rapid decline in abstract intelligence, spatial and physical abilities than in educational abilities was confirmed, and considerable evidence was obtained of greater declines among adults of initially low ability. Not only education but also intelligence appears to be better retained by those who "use their brains" more. Again, mechanical-spatial abilities show increases from 14 to 18 years; general intelligence probably rises only among those receiving further education, and educational abilities decline among those not receiving schooling.

Small but consistent geographical differences in intelligence were found between the east or south and areas which contain considerable proportions of Welsh and Irish recruits. Intakes into the three men's Services varied at different times, and average differences between Army, Navy and R.A.F. ground personnel were small. However, the Army certainly received the biggest proportions of low grade, and R.A.F. air crew large proportions of

high grade, recruits. Members of pre-Service organisations (A.T.C., Sea Cadets, Army Cadets, etc.), also Scouts, showed superior qualities in the Services. But this was attributable more to the intellectual, educational and other traits of youths who join these organisations than to the training as such, which was shown to produce rather limited effects. Physical training courses were found to improve the scores of Army recruits with poor physique on certain physical and mental tests, but menstruation among A.T.S. recruits had no consistent influence on test performance.

A very large proportion of recruits in the Forces and of adults in industry have not met an intelligence, or other psychological, test before, hence their scores are likely to improve considerably as they gain familiarity with such tests. It is important to determine the extent of such improvement, both because different recruits no doubt possess different degrees of previous experience, and because re-testing is frequently necessary when original scores, cannot be traced. We will first outline some of the published evidence on practice effects.

Dearborn and Rothney (1941) summarise most of the American work and show, from their Harvard Growth Study, that practice effects are generally not very large, but that they do occur with some—not all—group tests. They do not seem to be confined to any particular type of test material. With repeated testing the effects tend to diminish, i.e. the greatest increase is from the first to the second test. These authors also claim that practice on one test affects that test only and does not extend to other slightly different tests.

Some, but not all of these results are confirmed by British investigations, and we would suggest that the difference lies in the greater degree of "test-sophistication" among American children and adults. Not only do they habitually take more intelligence tests, but also most of their examinations are new-type ones, which are made up of questions similar to those in intelligence tests. We would therefore expect rises to be larger in this country, and spreading to occur from any one test to other dissimilar ones. For example, Vernon (1938b) investigated test-sophistication among university-trained adults and showed that the study of tests in general and their make-up, and practice in taking various types of test, raised scores both on verbal and non-verbal intelligence tests *which had not been taken before* by as much as 8·4 per cent.* An American study by Bowers and Woods (1941) points in the same direction. They compared the intelligence test scores of over 1,700 students who had taken none, one, two and three or more intelligence tests during their school careers. If the scores of the no-experience group are reduced to 100, the scores of the other three groups average 106, 108 and 110 respectively. Evidence is then given to show that the differences even up in the course of a year's work at college, and that they are far greater among students from rural areas and small towns than among students from large towns. A plausible explanation for this is that the big high schools in large towns would make more use of new-type achievement tests, and that their ex-pupils would be more test-sophisticated.

A. G. Rodger (1936) applied six parallel tests at fortnightly intervals to ninety-five British 11-12 year old pupils and claims that the average increase was about 1 per cent., or 1 I.Q. point, per test. His figures suggest, however, a rise of 3.8 per cent. between the first and third tests and thereafter no further change. The Moray House tests which he used have a fore-exercise or practice sheet, which probably minimises the susceptibility of the tests themselves to practice. Rodger also states that the rise was greatest (1½ points) in brighter children of I.Q. about 120, and lower (½ point) in dull children of I.Q. about 80. Dearborn and Rothney make a similar claim. It is a nice theory that the most intelligent, because of their intelligence, improve most. But all our evidence, cited below, shows the greatest rises among those scoring least. Dearborn's and Rodger's conclusion is, in fact, based on changes in test norms, not on rises among actual pupils.

McRae (1942) likewise gave sets of six tests to small groups at weekly intervals and confirms the diminishing effects of repeated practice. He concludes that when testees vary initially in their previous familiarity with tests, a single test will act as a sufficient "shock-absorber" to bring them all on a par. He also noted, when giving parallel versions of a test, that the effects on Form B of actual coaching on Form A were no greater than the effects of merely doing Form A in the ordinary way.

Dave (1938) found that some types of test are much more affected by practice than others, non-verbal and spatial items apparently being more susceptible than most verbal items. Finally, recent researches by Heim, Wallace and Carpenter showed that continued practice on a single test may improve scores almost indefinitely*. When nine W.E.A. students took the A.H.4 test eight times in successive weeks, there was some tendency to flattening out after about the fifth occasion, but some subjects continued to rise until they attained nearly perfect scores. Nine seamen tested nine times on almost consecutive days gave similar results. These show the dangers of leakages of test material and of unauthorised coaching. It was noteworthy, however, that the correlations between "unpractised" and "practised" test scores remained extremely high. Thus a test may still be a valid test of intelligence after practice, provided that all testees have had the same amount of practice.

---

* All score changes quoted in this Chapter have been converted into a common scale, roughly parallel to intelligence quotients. That is, each alteration has been divided by the Standard Deviation of scores and multiplied by 15.

Turning to work in the Forces, the Matrices test was re-taken by 537 seamen in an entry establishment one to six months after it had been done at recruiting centres. The average rise was 4·7 points or 8·6 per cent. But as the reliability of this test is rather low, the correlation between the two sets of scores being only ·79, the alterations were irregular. Some men actually declined on the second occasion, and the total range of changes was from 25 points increase to 13 points decrease. A natural consequence of what is called the regression effect is that very high scorers showed least improvement, very low scorers most.

An experiment in the Army where the same test was re-taken by 277 men after only one day yielded almost the same rise, and other later work suggested that practice effects are much the same after several months as they are immediately. But a possible alternative explanation is that military or naval training, including the taking of proficiency examinations, also helps performance at tests, hence a rise after six months may be partly due to this and only partly to recollections of the previous testing. Sometimes, of course, the training received is directly relevant to the abilities tested, e.g. mathematical or mechanical. This would account for the different findings in the major Naval and Army experiments on test reliabilities and re-test rises.

The Army results listed in Table XIII were based on re-testing

---

* From unpublished work communicated by courtesy of the M.R.C.'s Unit for Research in Applied Psychology, at the Cambridge Psychological Laboratory.

TABLE XIII.—PERCENTAGE INCREASES ON RE-TESTING AMONG ARMY AND NAVAL RECRUITS

Test	*Percentage Rises*	
	Army	*Navy*
1 Abstraction	—	3·2
2 Bennett Mechanical	4·6	10·4
3 Arithmetic Mathematics	2·1	8·6
4 Squares	8·0	10·6
25 Verbal	3·3	—
12 Clerical	6·6	—
8 Assembly	6·0	—
10 Morse Aptitude	3·9	—
16 Agility*	7·2	—

500 men, representative of the total intake, after eight weeks of primary training (which involved little or no bookwork). In the Navy, however, 500 air mechanics were re-tested after six to eight months of mechanical training. In spite of the longer interval their increases on Tests 2 and 3 are far larger. Squares (Test 4) is slightly larger, but Abstraction (Test 1, not taken in the Army) was probably uninfluenced by their training, hence it shows only a small rise.

The usual regression effect was observed on most tests. The top 10 per cent. of scorers achieved only a slight improvement or even declined on the second test, and lower groups showed successively greater increases. But in three tests—Verbal, Clerical and Squares (the latter in both Services)—men scoring in the 90th to 50th percentiles at the first test showed as great improvement as those in the bottom half. Probably this is due to the positively skewed or almost rectangular distributions of scores on these tests.

The total increase on T2 in the Naval group was larger than that on any component test, namely, 12·4 per cent. This is very considerable, being equivalent to an average rise of nearly 1 S.G. Even if half of it is attributable specifically to the mechanical training, it means that T2 is very sensitive to the practice effects of previous testing and of more generalised schooling.

Some evidence indicates that the effects of taking a different set of tests before are smaller than the effects of doing the same test. The average increase, shown in Table XIII on eight Army tests, also on Matrices and Abstraction, due to taking the same test before (along with others) amounts to 5·3 per cent. But in an experiment described below (p. 197) the Naval T2 tests and the R.A.F.'s G.V.K. tests were taken by the same recruits a few days apart, and the practice effect on whichever battery was taken second was estimated as 3·6 per cent. This is similar to the 2½ per cent. (2 to 3 points of I.Q.) admitted by Terman and Merrill when Form M of the revised Stanford-Binet test is given shortly after Form L, or vice versa.

---

* The high figure for Agility might be ascribed to improvement of physique among recruits in their first eight weeks of Army life. However, in an earlier experiment, an almost immediate re-test yielded a still greater improvement of nearly 12 per cent.

Further consideration of Table XIII suggests that the biggest effects occur in tests which are most novel, or which have most complicated instructions. Squares and Clerical are particularly difficult to "get across." The Arithmetic test, which is entirely straightforward and familiar, has the smallest increase of 2·1 per cent., and the Naval Abstraction test, whose instructions are so simple that it is self-administering, is also very low. This is borne out by the following investigations.

Several versions of a mechanical comprehension and information test, and of a mathematics test, were devised for use at recruiting centres, and were simplified in order to make them self-administering. These were tried out on 1,400 recruits in H.M.S. *Royal Arthur* of whom 190 had taken a version of these tests a few months earlier at recruiting centres, while the remainder had taken Matrices. The former group showed no superiority on the mathematics section, and were only 1·5 per cent. superior on the mechanical section as a result of their previous experience.

In an Army experiment, Matrices was given twice on consecutive days with a forty-minute time limit to 270 recruits, and an alteration of under 6 per cent. was found, contrasting with the figure of over 8 per cent. obtained with a twenty-minute limit. The longer limit presumably allowed the testees to get more used to the unfamiliar task*. Other studies were made of dictation and spelling tests. Five passages were tried out on 647 seamen in an entry establishment, each class getting a different set of two passages in different order. The average scores on the second passage were only 1·2 per cent. superior, and this is perhaps attributable to familiarisation with the dictator's pronunciation. In another research, four groups of fifty recruits took five tests in immediate succession. One was the A.T.S. Spelling, two involved dictation of sentences in which one difficult word had to be written down, and two were of a novel type. Each item consisted of three near-synonyms, one of which was incorrectly spelled. Testees had to identify and re-write the wrong words. For example:

---

* In this instance the "ceiling effect" may have entered; with a 40-minute limit scores tend to be so high that little improvement is possible among the brighter testees. On most of the tests whose results are quoted in this chapter, however, there was ample "ceiling."

STREET RODE AVENUE.

No practice effect whatever could be discovered in any of the tests except this last, unfamiliar, one where it amounted to about 2 per cent.

The following appear to be the main conclusions arising out of the investigations we have described:

(1) Test scores are seriously disturbed if different testees have had different amounts of practice on the particular test.
(2) The effects of taking other similar tests are smaller though still very considerable.
(3) Such practice tends to show diminishing returns.
(4) The effects of practice are probably lasting, being almost as great after some months as after a few days. This point requires much fuller investigation.
(5) Apart from training in the subject matter of the test, it is possible that any kind of schooling and the taking of examinations (especially new-type ones) have appreciable effects.
(6) The effects tend to be greatest among those who are least accustomed to tests, and possibly to bookwork in general.
(7) The effects are much smaller in straightforward inventive-response tests with simple instructions and ample time allowance than in unfamiliar tests with elaborate instructions, also in verbal than in most non-verbal and spatial tests.
(8) They can be reduced by adequate fore-exercises. Moreover, in view of No. 3, testees with different amounts of experience can be brought to a more even level by taking one or two preliminary tests.
(9) Re-testing with the same, or parallel, tests is undesirable, but when unavoidable distinct norms or standards should be provided. So far as is known re-test scores have the same validity as original scores, but this, too, requires further study.

Age and Other Differences

Many results of general interest might have been anticipated from the large-scale testing of recruits during the war, for example, comparisons between the three Services, between men from different parts of the country, or from different occupations, and so forth. Such enquiries are, however, much more complicated than they appear at first sight since age, occupation, recruitment policy and other factors that influence test scores are so interwoven. Eighteen-year-old recruits were probably fairly representative of the population as a whole, at least in 1942–3, though even at this age a considerable number had their call-up postponed, since they were in reserved occupations; and these were mostly in skilled trades or professions and, therefore, of superior intelligence. A very small proportion of men of low medical category (mostly below average in intelligence), of psychiatrically unfit and of mental defectives, were rejected. The composition of older batches of recruits was extremely variable owing to the Ministry of Labour's reservation policy. In one month, for example, large numbers of 20–25 year old policemen were de-reserved and intakes into all the Services were of outstandingly high quality. In another month there was a big proportion of 30–40 year builders, and the quality of intakes was lowered partly because the average for building tradesmen and labourers is a little below that of the general population, partly because test scores tend to decrease with age. Again when an occupational group was partially de-reserved, employers would naturally tend to hang on to their best and presumably most intelligent men and to release the duller ones for the Forces. Thus observed age differences were distorted by differences in the occupational make-up of groups of recruits of different ages, and occupational differences were distorted by the age groups that happened to be called up during a certain period. Psychologists were, of course, unable to control these factors, or to hold them constant, as they would do in a laboratory experiment, hence it was almost impossible to disentangle their effects.

Pre-war research, though often on a small scale, had established the following facts by taking special precautions to secure representative cross-sections of the population:

(1) General intelligence increases on the average at a steady rate up to about 12 years. The rate of increase then declines until a maximum is reached by about 15 years, though on some

tests there is little if any increase after 13 and on others rises have been reported even beyond 18–20*.

(2) Between the twenties and sixties there is a progressive decline on tests of *g* involving abstract reasoning and speed of mental manipulation, though on other tests of what has been called "crystallised" intelligence, such as vocabulary and information, the level is better maintained (cf. Cattell, 1943; Brody, 1944).

(3) No one individual necessarily conforms closely to these average trends. "Longitudinal" studies of particular children often show great irregularities of mental growth, with spurts and plateaux attributable partly to emotional adjustment or maladjustment, partly to the stimulating or inhibiting effects of the child's home and school environment (cf. Fleming, 1948). The same may well be true of adults, but longitudinal studies are much more difficult since an adult cannot be re-tested many times without his scores being affected by "sophistication," and by his attitudes towards the investigation, e.g. growing hostility.

(4) Educational attainments tend to be forgotten rapidly after leaving school, except in so far as they are used in daily life. Thus, Norris (1940) finds that scores on linguistic tests may rise till about 40 years, but arithmetic achievement declines in most persons other than clerks who practise arithmetic in their job. The performance of adults on intelligence tests, according to Lorge (1945), is also affected by education beyond 14 years, though, as Garrett (1946) points out, results on a verbal test do not necessarily mean that intelligence itself alters. Adults with a university education, tested at 34 years, were about two years superior in mental age to others *of the same intelligence level when aged* 14 *years* who had received no secondary or university schooling. An investigation by Miles (1932) also appeared to show an earlier decline in intelligence among adults who had only had elementary schooling, and Cattell (1943) claims that superior occupational groups retain their intelligence better than lower ones.

Several studies in the Forces helped to supplement these conclusions. The abstract reasoning ability which is particularly affected by age is typified in the matrices test. Thus, it was somewhat unfortunate that this test should be the one most widely applied in the Royal Navy, Army and A.T.S. When it was first introduced the Navy happened to be recruiting men mostly aged 30 and over, and the norms which were established then, and which could not readily be altered, were extremely faulty when applied to later intakes consisting mostly of 18-year-olds. The S.G.s were chosen to divide the older population into 10, 20, 40, 20 and 10 per cent. groups, but with more representative recruits they yielded 12 to 14 per cent. of S.G.1s and only 1 to 5 per cent. of S.G.5s (cf. Table XIX)*.

---

* Dearborn and Rothney (1941) claim increases up to much later ages than 15, but they are assuming an indefinite continuation of schooling.

A typical set of correlations for nine tests with age was obtained in 1942 in a group of 578 Army recruits well spaced out from 18 to 40 years; these are listed in Table XIV. All of them, it will be seen, are negative, even with tests such as arithmetic and messages which involve most education. This suggests that the older age groups were of inferior quality apart from age. But the relative order of correlation is none the less interesting, showing that such "*v : ed*" tests are least affected, Matrices and tests dependent on sensory and physical qualities most affected (the Assembly test being exceptional possibly because engineering experience tends to increase with age). It was also found that the decline was fairly small until 35 years or more. When all men of 21 and over were contrasted with the 20s and under the mean (tetrachoric) correla-

TABLE XIV.—CORRELATIONS BETWEEN TEST SCORES AND AGE AMONG ARMY RECRUITS

Test	*Correlation*
Progressive Matrices	−·340
10 Morse Aptitude	−·296
16 Agility	−·208
4 Squares	−·193
2 Bennett Mechanical	−·182
12 Clerical	−·173
3 Arithmetic	−·165
8 Assembly	−·151
17 Messages (Verbal)	−·126

---

* It is an interesting point that the S.G.1s in the young samples have never reached as large proportions as would be expected from their superior mean scores. This indicates that, although the proportion of S.G.4s and 5s increases rapidly with age from 18 to 30+, the proportion of S.G.1s only sinks very slowly, so confirming the conclusions based on Table XVI on page 192.

tion for all tests was —·137. When the 30s and over were contrasted with 29s and under it was —·149, but when 35s and over were contrasted with 34s and under the average coefficient rose to —·288.

The A.T.S. differed from the Army in that recruits of, say, 25 and over were often of better quality, drawn from higher-grade occupations, than the younger ones, since more of them were volunteers. All their correlations with age, therefore, tended to be more highly positive or less highly negative than those of men. In one representative sample of 200, they ranged from —·100 for Matrices and —·070 for Squares to +·147 for Verbal (Test 25), and +·180 for Spelling. This —·100 for Matrices is almost as certainly too small, as the —·340 (in Table XIV) is too high. Hence the correlation of —·238 obtained among 90,000 naval recruits, whose occupational make-up was held constant, may well be about right (cf. Table XV).

Differential Decline with Age

Since age changes differ on different tests it should be pointed out that any composite intelligence score, based on several tests of varied types (such as T2 or Summed S.G.) will have a somewhat different make-up at different ages. A man of 20 and another of 35 may obtain the same Summed S.G., but the former is likely to score better on Matrices and Bennett, the latter on Verbal and Arithmetic.

That the rate of decline in *g* (as measured by Matrices, not by verbal intelligence tests) is greater at lower intelligence levels was indicated by a survey of 90,000 naval recruits in 1942 (cf. Vernon, 1947c). They were classified under twelve broad occupational headings, so that it was possible to maintain the same occupational distribution at all ages. Table XV shows the names of the groups and total numbers, together with the mean scores of 16- to 19-year-olds (average about 18·0) and the mean scores of 20- to 40-year-olds (average about 30·0). The last two columns show the decreases between the 16–17 and the 18–19 year groups, also between the 16–19 year groups and those aged 30 and over. In the first of the latter columns the decline in the four most intelligent occupations is less than a quarter the decline in the three dullest occupations. In the last column the results are more irregular, possibly because of the small numbers available at later ages, nevertheless, the

TABLE XV.—AVERAGE MATRICES SCORES AND DECLINES WITH AGE IN TWELVE OCCUPATIONAL GROUPS

Occupation	*No.*	*Mean Scores at*		*Decreases from*	
		18.0	30.0	16+ to 18+	16+ to 30+
Clerks	11,603	43·4	39·1	0·39	5·07
Electrical Workers	4,602	40·3	37·3	0·16	5·18
Woodworkers	6,705	39·0	36·2	0·25	3·40
Precision Workers	11,067	38·9	35·8	0·33	5·43
Sheet Metal Workers	6,401	37·4	35·1	0·49	3·72
Retail Tradesmen	9,373	37·4	34·7	0·77	3·80
Machine Operators	5,936	36·7	33·4	0·42	5·67
Builders	7,074	36·4	32·9	0·77	4·43
Drivers	5,764	34·6	31·6	0·51	4·66
Mates	5,367	36·6	30·6	0·85	8·49
Farm Workers	2,406	33·4	30·5	1·39	5·17
Labourers	13,466	33·6	28·6	1·57	6·99
All	89,764	37·4	34·2	0·72	5·26

decline is more than twice as great among labourers and mates as among woodworkers and sheet metal workers. The interpretation of these figures is dubious both because the 16–17-year-olds consist of volunteers for the Navy, whereas the 18-year and later groups consist mostly of less intelligent conscripts, also because of unknown effects of de-reservation policy. It might be rash to conclude that the general decline in *g* starts as early as 17. There is, however, no doubt of the statistical significance of differential rates of decline in different occupations, and they appear to fit the hypothesis of more rapid decline among men who make least "use of their brains."

The same point is brought out in another way in Table XVI, which shows the percentages of successive age groups obtaining very high Matrices scores—50–54 and 55–60, together with the

TABLE XVI.—PERCENTAGE MATRICES DISTRIBUTIONS AND MEAN SCORES OF DIFFERENT AGE GROUPS

Score	*Age (years)*						
	16-17	18-19	20-24	25-29	30-34	35-39	40+
55-60	0·33	0·36	0·32	0·35	0·15	0·04	0·00
50-54	3·72	3·71	4·02	2·55	1·69	1·29	0·69
35-49	65·85	62·64	59·30	52·99	48·50	42·80	33·13
0-34	30·10	33·29	36·36	44·11	49·66	55·87	66·18
Nos.	32,412	30,451	6,165	4,060	5,351	7,577	3,748
Means	37·68	37·13	36·43	34·66	33·74	32·40	30·30

numbers and mean scores for all occupations combined. In spite of the steady decrease in the mean with age and in the proportion of above average scorers (from 70 per cent. at 16–17 to 34 per cent. at 40 +), the proportion scoring 55 and over remains steady (or even rises) till 29 years, and only then starts to drop rapidly. Again the proportion scoring 50–54 remains steady (or rises) to 24 years, and then drops. While there may be some alternative explanation, it certainly looks as if the most intelligent retain their intelligence longest.

Changes with Age from 14 to 18+

In an investigation of naval artificer training establishments, where boys enter at 14 and are trained as skilled tradesmen till 18, S.P. Tests 1–4 and several supplementary tests were given to all boys, including 300 of average age 15¼ and 250 of average age 18¼. Although this population is far from representative of the general average, being highly selected educationally, it has the advantage of remaining stable in composition throughout the four years since (unlike the ordinary secondary school) scarcely any boys are eliminated once they have started the course. The gains in scores from 15 to 18 listed in Table XVII have been converted into a percentage scale to make them comparable for all tests. There is a considerable improvement throughout, and it is almost as great in the purest *g* test (Test 1) as in mathematics—that is a school subject in which boys receive a fair amount of instruction. As one would expect it is greatest in the information tests (117E, M, and 2) on

TABLE XVII.—PERCENTAGE INCREASES OR DECREASES IN TEST SCORES BETWEEN 14+ AND 18+ AMONG ARTIFICER APPRENTICES AND AMONG BOYS LEAVING SCHOOL AT 14

Test		*Gains or Losses Among*	
		Artificers	*14-yr. Leavers*
1	Abstraction	+4·8	+1·0
2	Bennett Mechanical	+10·0	+5·7
3a	Arithmetic	+5·7	−4·2
3b	Mathematics	+6·5	−9·0
4	Squares	+9·9	+5·3
	T2	+12·4	+1·9
97	Memory for Designs	+5·5	—
117	Mechanical Information	+14·2	—
	Electrical Information	+21·6	—
110	Mechanical Models	+4·0	—

account of the training, but it is very large also in Test 4, that is in spatial ability which is not directly trained at all.

During 1946–7 several S.P. tests were applied by one of the writers and his students at the University of Glasgow to large groups of school children. One sample* included all the available 14-year-old boys leaving school in January, 1947, totalling some 1,200. Naturally this was a rather inferior sample, but their scores on the T2 tests were compared, not with the standards for naval recruits in general, but with the scores of a group of 265 naval recruits who left Glasgow schools at the age of 14 in 1938 and who entered the Navy in 1942. The percentage increases or decreases in this group over the 14-year-boys also appear in Table XVII and are very different from those found among boys of superior intelligence whose education was continued. There is only a negligible increase in the all-round ability measured by T2 or in the *g* measured by Test 1. On educational tests there is a serious loss. On the other hand there is a considerable rise in mechanical and spatial ability, although it is unlikely that many of the 18-year recruits had received any systematic technical or trade training. It would appear then that *k : m* abilities rise with or without schooling from 14 to 18, that *v : ed* abilities drop and *g* remains fairly constant, unless they are stimulated by secondary schooling. But the constancy in all-round ability measured by a battery such as T2 is specious, and occurs merely because the gain on the practical side roughly balances the loss on the educational side.

In other post-war enquiries, approximate age norms were established for a version of Test 1 (Abstraction), for a spelling test, for Army Arithmetic and for the Kohs-Misselbrook blocks test. The average adult standards, as obtained in the Forces, were found to be reached by children of 13, 12½, 13 and 15½ years respectively. That is to say, the average 18-year recruit may actually be lower on the first three, and particularly on spelling, than he was at the age of 14, but—just as was observed with Test 4—he is better than the 14-year-old on a performance test which involves *k*-factor. Further analysis of the Abstraction, Spelling and Arithmetic results showed that the standards at 14 and 18 are just about the same at the top end of the scale. The best 10 per cent. of adult recruits do as well as the best 10 per cent. of normal 14-year-olds. But there is a progressively greater decline at the lower end. The bottom 10 per cent. appear to lose 1½ years on Abstraction, two years on Arithmetic and three years on Spelling. At 14 years the 10 per cent. poorest in spelling reached a maximum level equivalent to the average 11-year-old; whereas among recruits the corresponding level is only 8 years. It is little wonder that there were so many complaints of semi-illiteracy and appalling spelling among recruits when dull adolescents lose so rapidly the skills in which they were drilled till 14. The better retention among brighter men, who receive further schooling or who use their skills in their jobs, confirms our conclusion as to the differential decline in abilities.

---

* This testing was part of a survey undertaken by the Social and Economic Research Department of Glasgow University. The boys were scattered over 60 schools, and the writer is grateful to Miss C. McCallum and the staff of the Glasgow Corporation Child Guidance Clinics for arranging and carrying out the testing and scoring.

We see then that the average performance of a group of adolescents or adults on psychological tests varies greatly with the type of ability tested, with age, and with any schooling or other training the group has received, or forgotten. This complicates tremendously the establishment of satisfactory test norms. A further corollary, since tests of *g* appear to be affected in much the same way as educational knowledge by use or disuse, is that psychologists and educationists should investigate just what types of adolescent and adult education, and of occupational and avocational experience, most effectively stimulate intelligence. Not only can it be raised during the 'teens by schooling, but also the inevitable decline in adulthood can probably be retarded.

Occupational and Geographical Differences

The striking differences on T2 and other tests between men in various Service employments are described in the next Chapter. Few large-scale studies of civilian occupational differences were carried out, apart from the one whose results are listed in Table XV. This was the only one, too, where the age distribution was held constant in all groups. It may be observed that the range of scores from clerks at the top to labourers at the bottom is rather small, corresponding in terms of I.Q. to a range of about 110 to 90*. Verbal tests such as Army Alpha and Cattell's Scale III give a much wider range than does Matrices. This bears out our contention (in the next Chapter) that occupational suitability, and, therefore, occupational level, depends on *v : ed* factor as well as on *g*.

---

* This assumes a Standard Deviation of 16. Cattell's Scale III has a S.D. of 25. Had we assumed this figure our range would be perhaps 116 to 84. But Cattell's range is still much wider, running from clerks with I.Q. 127 to packers and sorters I.Q. 78.

The only regions of Britain studied were the nine very heterogeneous areas into which naval recruiting centres are grouped. Small yet significant differences are indicated by the mean scores listed in Table XVIII. The first column of means gives the obtained figures, corrected only for age differences. But occupation and area are closely linked. Glasgow is low partly because it contains the largest proportion of labourers, Manchester high because it has an excess of clerks and electricians. In the last column, therefore, the occupational distribution has been equalised in all areas. Manchester then drops a little and Glasgow rises, but the alterations are only slight.

TABLE XVIII.—MEAN MATRICES SCORES OF NAVAL RECRUITS FROM DIFFERENT AREAS

Area	No.	Mean	Scores
		(i)	(ii)
London, including Kent, E. Anglia and Northampton	22,376	37·5	37·3
Manchester, including Yorkshire	12,625	37·2	36·7
Derby, including Lincoln and Nottingham	5,496	36·5	36·5
Southampton, including Sussex, Oxford, and Dorset	7,705	36·3	36·8
Newcastle, including Cumberland, Westmorland and Durham	5,264	36·0	36·2
Liverpool, including W. Lancashire and N. Wales	7,690	35·8	36·0
Birmingham, including Gloucester, Hereford, Worcester, Shropshire and Mid-Wales	9,397	35·5	35·5
Bristol, including Wilts, Somerset, Devon, Cornwall and S. Wales	7,988	35·0	35·2
Glasgow, including the whole of Scotland	11,223	34·8	35·3

Naturally a bigger range of differences would be expected between more homogeneous areas. It is absurd to group, e.g. Scottish Highlanders with Glasgow Irish, and to make no separation between predominantly urban and rural regions. Nevertheless, even these crude figures are of interest in suggesting national differences, though considerable caution is necessary since we do not know how representative are the samples. Welsh and Irish recruits would occur chiefly in Liverpool, Birmingham, Bristol and Glasgow areas, and it is noticeable that these are at the bottom of the list.

Another comparison was afforded by the survey of Glasgow 14-

year school leavers, though these groups are not, of course, typical of the whole 14-year population. The mean T2 for 400 boys at Roman Catholic schools, i.e. mostly of Irish descent, was 55·0, and for 771 boys in other schools 64·3. This corresponds to a difference in average I.Q. of about 7 points.

Service Differences

The belief was widely held that the R.A.F. got most of the "cream" of recruits, that the Navy also had more than its share, while the Army had to put up chiefly with below-average men. This could not be directly proved or disproved since no one test was taken by all three Services. However, a fairly trustworthy "bridge" was built up as follows. The G.V.K. tests were given to a group of 552 naval recruits, half before and half after taking the T2 tests. (The two batteries gave an inter-correlation of ·792.) Allowance could thus be made for practice effects, and corresponding percentile levels on the batteries were found. Another conversion table between Matrices and T2 was constructed (their inter-correlation in an unselected group of naval candidates being estimated as ·80). Now Matrices had been given to all Army and A.T.S. recruits in 1942 as well as in the Navy. Table XIX shows the percentage S.G. distribution on Matrices for some 100,000 Army recruits entering during July–November, 1942, for an equally large group of accepted naval candidates during January–September, 1942, for 3,759 ordinary seamen within the same period, and for A.T.S. intakes during the whole of the same year*. It will be seen that the intakes are closely similar in numbers

TABLE XIX.—MATRICES DISTRIBUTIONS DURING 1942 IN THE ARMY, NAVY AND A.T.S.

S.G.	*Score*	*Army*	*Naval Acceptances*	*Seamen*	*A.T.S.*
		per cent.	*per cent.*	*per cent.*	*per cent.*
1	46+	12·3	14·5	11·9	14·3
2	40+	23·6	27·9	30·6	27·6
3	29+	42·5	44·1	40·8	39·3
4	20+	16·2	12·3	14·0	17·5
5	19−	5·4	1·2	2·7	1·3

---

* These are Army norms; unfortunately naval and A.T.S. norms differed by 1 or 2 points. Since S.G.s alone were recorded at Recruiting Centres, not scores, only the Army and the O.S. figures are the percentages actually observed. The naval and A.T.S. figures have been converted to Army norms, and are unlikely to be inaccurate by more than 0.5 per cent.

of S.G.1s, but that it is true that the Army received a much larger proportion of S.G.5s or very dull recruits. In terms of scores the mean difference between Army and Navy amounted at most to 2 points on Matrices at this period, or 5 points on T2.

We can probably allow that the Army intakes were slightly inferior to the population as a whole, and that naval and A.T.S. ones were slightly superior. Naval norms were, however, mostly based on ordinary seamen, not on recruits in general. This meant that some of the brightest recruits, entered as radio mechanics, telegraphists, etc., and some of the dullest entered as stokers and cooks, were omitted. Table XIX shows that seamen were of slightly lower quality than naval acceptances, though better than Army recruits, and it is a reasonable guess that, in average score at least, they were very close to the norm for the population as a whole. Probably, however, they were more restricted in range than the general population, both because very low Matrices scorers were rejected at recruiting centres, and because of the omissions just mentioned.

Table XX gives the best estimates that could be made of T2 percentiles in R.A.F., Navy and Army. Being based on moderate-sized groups they are less representative than the figures in the previous Table, and the 90th and 10th percentiles are naturally less reliable than the 50th. The table bears out our conclusion that in 1942 Army recruits were only slightly poorer than naval ones, and it may be seen that R.A.F. ground staff were but little superior to seamen. By 1944–5, however, the Navy was able to reject many more candidates at recruiting centres, whereas the Army could only raise its standards to a small extent. These figures remain fairly representative of peace-time intakes. R.A.F. air crew, in

TABLE XX.—T2 LEVELS IN R.A.F., NAVY AND ARMY

Service	*Date*	*No.*	*Percentiles*		
			90th	*Median*	*10th*
R.A.F. Air Crew	1942	1,141	141	114	85
R.A.F. Ground Staff ..	1942	5,000	116	81	54
Navy O.S.	1942	1,000	108	75	46
Navy O.S.	1944	3,384	132	100	68
Army recruits	1942	578	100	72	36
Army recruits	1945	1,000	112	81	48

contrast with ground staffs, were clearly very superior, and provided some substance for the Army's complaint that the R.A.F. and Navy got the largest share of high-grade material.

Effects of Pre-Service Training

Great stress was laid by the Services on the value of pre-Service training in the Sea Cadet Corps, Junior or Senior Training Corps or Army Cadet Force and the Air Training Corps. Several follow-up studies in the Royal Navy showed that such recruits, together with members of Scouts and Boys' Brigade—who constituted a quarter to a third of all intakes—were superior as seamen, in Coastal Forces, as radio and air mechanics and air fitters, and as telegraphists; but it was doubtful how far this was due to their possessing better *g* and education. Some 4,500 recruits in H.M.S. *Royal Arthur* were studied in 1944, including nearly a thousand Sea Cadets and a thousand A.T.C. members, and smaller groups from other organisations. The following rank order of average ability was obtained on T2 and on Test 3b (Mathematics) and educational level:

J.T.C. and S.T.C.
Scouts and Sea Scouts.
A.T.C.
Members of no organisation.
Sea Cadets.
Boys' Brigade and Church Lads' Brigade.
Army Cadets.

It should be remembered that the best A.T.C. members were likely to join the R.A.F. and the best Army Cadets the Army, also that the general quality of entries was high at this time. In 1942–3 Sea Cadets would certainly have been superior to average intakes. Tabulations were made of the number of recruits in each group recommended by P.S.O.s as suitable for the main categories—officer candidates, mechanics, writers and coders, communications, seamen and stokers, cooks and stewards, etc. It was found that Sea Cadets provided rather more trainees for the most active branches—namely, officer candidates and seamen, though fewer for the specialist branches than did non-members, even when allowance was made for lower intelligence and education. The A.T.C., when ability was held constant, provided an excess only in the communications branches. Ex-Scouts yielded at least as large a proportion of officer candidates and specialists as any of the pre-Service organisations. Length of membership of the organisations was found to have a favourable, though only a very slight, effect on the yield of high-grade material. The conclusion reached was that pre-Service organisation membership as such had scarcely any effect, except for morse and mathematical training given in the A.T.C. The good showing of members is almost entirely attributable to the intellectual, educational and other qualities of those who join the organisations.

Similar results were obtained in the Army when some 3,500 members of pre-Service organisations, and 5,500 non-members were studied. On test results, education, and ratings by P.S.O.s for officer quality and leadership, the rank order of organisations was Junior Training Corps, A.T.C., Army Cadet Force, non-members, and Sea Cadets. The latter now come at the bottom since the best cadets naturally prefer the Navy. While 12½ per cent. of members and only 2½ per cent. of non-members were earmarked as potential officers, the difference was almost wholly attributable to the higher education and intelligence of the former—as shown by Table XXI.

TABLE XXI.—ANALYSIS OF VARIANCE IN POTENTIAL OFFICER RECOMMENDATIONS ATTRIBUTABLE TO SEVERAL FACTORS

	Sum of squares per cent.
Differences in educational standard	19·88
Additional variance due to membership of a pre-Service organisation	0.95
Additional variance between different organisations	3·14
Individual differences (residual)	76·03

Follow-up to later stages in the naval or military careers of these recruits would be desirable, but was not found practicable. In a smaller, but carefully controlled, experiment, fifty-five Army Cadets who had obtained the War Certificate A for proficiency were matched with non-members of equivalent education and intelligence, and were subjected to special training. Of the former group 83 per cent., of the latter 50 per cent., were able to accomplish their primary Army training in four weeks instead of the usual six, and were superior also in Corps training which was reduced from ten to six weeks. The Certificate recruits were better in rifle shooting, but showed little difference on light machine-gun, i.e. a comparatively unfamiliar weapon or on physical training.

Three months later over half the men had been before W.O.S.B.s, and the superiority of the Certificate holders was much less marked as shown by the correlations in Table XXII. On the other hand, the weighted selection test scores used for picking potential officers (the O.R.1 Index, cf. p. 46) correlated almost as well with W.O.S.B. as with primary training results.

TABLE XXII.—RELATIVE PREDICTIVE VALUE OF PRE-SERVICE TRAINING AND OF SELECTION TESTS

	Correlations with	
	Primary Training	*WOSB Pass or Fail*
Possession of Certificate A *v.* no pre-Service experience	+·47	+·24
Selection test battery (O.R.1)	+·38	+·33

Effects of Physical Condition

Some 250 Army recruits with very poor physique who took a two months' course at Army physical development centres were tested before and after. (Results on the Matrices test alone were available for 648 cases.) The gain in scores to be expected on re-testing was known from an earlier investigation into the reliability of S.P. tests. In every test there was a greater gain than mere practice effect, though it was often small and statistically non-significant. The results are expressed on a common percentage scale in Table XXIII. The increase in Agility would be expected and conforms with an improvement in medical category, height and weight. Attached to the centres are picked educational sergeants who organise numerous "brains trust" periods, spelling bees, and the like. This may account for the rise on the Verbal test. The increase on Matrices is less readily explicable, but suggests

TABLE XXIII.—PERCENTAGE GAIN IN TEST SCORES ATTRIBUTABLE TO ATTENDANCE AT P.D.C.S

Test	*Per cent. gain above practice effect*	*Probability of gain being due to chance*
16 Agility	6·5	< ·001
Matrices	2·8	"
25 Verbal	2·4	"
8 Assembly	2·2	< ·05 > ·01
3 Arithmetic	0·6	> ·05
2 Mechanical	0·5	"

that the test involves some kind of mental alertness linked to physical alertness, in addition to *g*. Note, however, that, although highly significant, it only amounts on the average to 1½ points of score.

Several enquiries were made in the A.T.S. into the effects of menstruation. The day within the menstrual cycle on which a battery of eight selection tests was taken for a second time was ascertained by a medical officer from 1,335 auxiliaries, all of whom claimed a normal or twenty-eight-day cycle (some 270 others admitting short, long or variable cycles were excluded). They were classified into four "phase" groups:

(i) From four days before to four days after the onset of menstruation.
(ii) From fifth to tenth day.
(iii) Ovulation phase—eleventh to eighteenth day.
(iv) From nineteenth to twenty-fourth day.

At the original testing the menstrual days were unknown and could be assumed to be randomly distributed. Comparisons of test and re-test Matrices scores yielded the analysis of variance shown in Table XXIV. There appeared to be slight differences between observed and expected re-test scores on some days, but they were irregular and were not associated with any particular menstrual phase. Phase as such had no demonstrable effect. A similar analysis of the Clerical test (S.P.12) gave negative results. Although very slightly lower scores were obtained on most of the eight tests by the period group (Phase i), in no case was it significant, even with this very large number of cases.

TABLE XXIV.—ANALYSIS OF VARIANCE DUE TO MENSTRUAL PHASE

	Sums of Squares per cent.	*P.*
Total Variance on Re-testing	100	—
Accounted for by regression of re-test on test	62·328	—
Accounted for by difference between phases	·066	> ·05
Accounted for by differences between days in the same phase	1·056	< ·05 > ·01

In a further investigation, 1,000 auxiliaries were asked to state both at test and re-test if they felt unable to do themselves justice, and if so on what grounds. The most common complaints were colds 10·4 per cent., menstruation 3·45 per cent., and chilblains

2·1 per cent. In no case did the scores of those claiming unfitness on either occasion show significant drops. Indeed, the most consistent difference was a rise in the practical Mec test scores (S.P. 24) among those undergoing menstruation.

Among some 3,000 auxiliaries enquiries were made at a medical interview and 24·4 per cent. reported some pain during their periods. A strong association was discovered between this incidence of pain and type of A.T.S. or civilian employment. It was significantly higher among women in strenuous or outdoor, than in sedentary or indoor, jobs. No significant relationship to educational standard was observed, and though pain occurred slightly more often among women of below average intelligence on Matrices, this was probably due to the association both of pain and of low intelligence with low medical category.

CHAPTER XII

GENERAL SURVEY OF THE VALUE OF VOCATIONAL TESTS

Abstract.—An outline of the main results of pre-war investigations indicates that tests of general intelligence, of mechanical, spatial and other abilities, are of considerable value in assessing occupational suitability. Mechanical-spatial tests are certainly superior to verbal and educational ones in the selection of adolescents for engineering training, though giving less consistent results with adults. Paper-and-pencil group tests of the miniature situation type and diagnostic tests show promise in a variety of educational and vocational fields. Though many psychologists consider that such general tests are less appropriate than specialised practical tests for particular jobs, the evidence is inconclusive, and the choice depends largely on whether the tests are to be used primarily for classification and guidance or for selection.

Very extensive follow-up investigations were carried out in the Services. These demonstrated first the value of a standard battery of group tests in a large variety of jobs, and secondly, the rather small extent to which more specialised tests helped in differentiating between different jobs. Mathematical and verbal tests tended to surpass mechanical and spatial tests even in mechanical and practical occupations. This tendency persisted when operational, as contrasted with training, criteria were studied, but would not necessarily be true under industrial—as distinct from Service—conditions.

So great is the volume of investigations into the value of different types of tests for predicting educational or vocational success, that we can attempt to give here only a very general summary of the literature before turning to the main results achieved in the Forces.

The usefulness of group intelligence tests of the ordinary verbal type in relation to school and university work is well established*, but it is generally considered that they have little bearing on vocations other than those which involve manipulation of words and numbers. Thus they give moderate correlations with proficiency among clerical workers, but have been shown in various studies to be unrelated to success among business executives, telegraphists, compositors, motormen, toolmakers, certain types of assembly workers, packers, etc. In jobs of a routine mechanical kind high intelligence and education may sometimes even be disadvantageous, leading to dissatisfaction and increased labour turnover. Not all these studies will bear close scrutiny, however; the tests used were sometimes inappropriate, the criteria far from reliable, and often no attention was paid to the selectivity of the groups concerned. Other investigations such as those with Army Alpha and Cattell's Scale III (1934), have shown very considerable differentiation in average intelligence between high- and low-grade occupations. Though there is much overlapping, for example, some machine operators being more intelligent than some schoolteachers, yet it must follow that intelligence has a bearing on success as a teacher. But as such groups as teachers are usually highly selected in respect of some factor like education, which is itself highly correlated with intelligence, any correlations between teaching ability and intelligence are thereby greatly reduced.

---

* A useful account of investigations at the University level is given by Eysenck (1947b).

Actually there is quite a lot of evidence of positive correlation between intelligence as measured by suitable tests, or educational standard shown by college grades, and success in professional and administrative jobs. Jenkins (1947) summarises studies of "leaders" in industry and other fields. The American Army General Classification Test gave moderate correlations with the achievement of commissions by officer candidates, though none at all with assessments of efficiency in battle. Barr (1946) and his collaborators found that when teaching efficiency is objectively measured by the progress made by the pupils, intelligence is one of the most predictive factors. In this country Heim (1947) has obtained promising results with her high-level A.H.5 test among industrial executives and B.B.C. engineers. Farmer and Chambers (1936) applied Group Test 33—the N.I.I.P. verbal intelligence test, the Cube Construction performance test, and various mechanical and hand-eye co-ordination tests to numerous groups of engineering workers. They obtained correlations up to about ·4 between the intelligence tests and proficiency in the more highly skilled, though not in the simpler, jobs. In Holliday's (1943) follow-up of engineering apprentices, Group Test 33 correlated as highly as his spatial and mechanical tests with instructors' assessments, though less well among trade apprentices. He suggests, however, that boys who do better on intelligence than on mechanical tests impress their assessors as being "bright," but do not make as good craftsmen. Both he and Farmer provide some evidence that tests are *more* predictive of proficiency after several years than they are of success in the early stages of training.

Nothing appears to have been published on the vocational value of non-verbal *g* tests. They are certainly less useful than verbal tests in relation to primary school work, but seem to be superior to them in predicting mathematical and scientific achievement at advanced secondary and university levels. Performance tests such as Minnesota, Kent-Shakow, Moorrees and Oakley Formboards, Cube Construction and O'Connor's Wiggly Blocks test are widely used in industry and in vocational guidance. There is, however, remarkably little evidence of their validity. In Rodger's (1937) study of several trades taught to Borstal youths, Cube Construction was the best test. Psychologists have largely employed such tests for the qualitative indications they yield of methods of work, but these, too, are in need of validation.

Mechanical assembly tests such as Stenquist's appear to have been used only on a small scale because of the time required for individual testing. No follow-up evidence is given on the 14,000 American Army recruits to whom an assembly test was applied in 1917–18. Fair results among boys and apprentices were found by the Minnesota investigators (Paterson *et al.*, (1930), Rodger (1937), Earle and Macrae (1929), Farmer and Chambers (1936) and others), and high validities have been claimed among certain types of metal workers and cotton-mill machine fixers. The Purdue Mechanical Assembly Test (Tiffin, 1946) is a new and improved version which has been validated among machinists. Group tests based on deductions about working mechanical models have been devised both by Cox and Vincent. These appear to be effective substitutes for assembly tests among adolescents, according to Hunt and Smith (1945), Holliday (1943), and Shuttleworth (1942), but they, too, have seldom been tried out on adult workers.

Group tests of spatial ability or *k* which have been applied vocationally include the N.I.I.P. Form Relations and Memory for

Designs Tests, the Minnesota Paper Formboard, Thurstone's space tests, Group Test 80, Squares and Figure Construction (cf. pp. 236–239). Many of the investigators just cited provide evidence of their usefulness in selection for technical education and trade apprenticeship. Certainly, along with models or assembly tests, they are superior to scholastic examinations, and probably to verbal intelligence tests, in picking boys for "practical" careers, though how far they work at as early an age as 11–12 is more doubtful. Holzinger and Swineford (1946) similarly report that they correlate with shop work and mechanical drawing, but not with geometry.

Several mechanical tests are based on pictures of machines, tools or mechanical situations. Cox's (1928) Diagrams, Explanations and Completion tests set out to measure comprehension or mechanical understanding rather than experience, while O'Rourke's so-called Mechanical Aptitude test (which also includes verbal items) is clearly a test of mechanical information. Stenquist's test with the same name and Bennett's Mechanical Comprehension test are intermediate, the latter also containing problems from statics and dynamics, heat, light and sound. There is considerable confusion as to what these tests measure, and their titles are often misleading. However, as pointed out above (p. 170), the distinction between aptitude and attainment is of little practical importance. Cox's tests are considered useful by the Birmingham investigators; O'Rourke's was widely used by the Tennessee Valley Authority, and the Bennett test was successful among machine-tool operators. But there is little further evidence of their validity from civilian investigations. Cunningham (1943), describing the war-time applications of psychological methods in Australia, mentions a study of twenty tests for selecting fitter and turner trainees in munitions industries. The most valuable were intelligence, technical information, and mathematics tests, Cox's Diagrams, a paper formboard and other *k* tests.

All the above tests are general ones—tests of general intelligence, spatial, mechanical or practical abilities. Most psychologists assume that in selection for particular jobs, tests must be more specifically designed on the basis of a careful job analysis. Drake (1940), for instance, as a result of extensive experience in a large industrial firm, regards the application of paper-and-pencil tests as a waste of time and money. He himself developed dexterity, co-ordination and other tests directly related to groups of similar jobs. On the other hand Pond (1941), in the course of many years' work in a metal manufacturing company employing 5,000 people, appeared to obtain good results with a verbal intelligence test, paper formboard, the MacQuarrie paper-and-pencil mechanical ability test, O'Connor Wiggly Blocks and Kent-Shakow formboard. Similarly the personnel department at Rowntrees uses a limited battery for all entrants.

Descriptions of typical analytic, analogous and work-sample tests are given in any book on industrial psychology. Surveys of recent work on selection tests have been published by Long and Lawshe (1947) and Hardtke (1945), the latter listing over a hundred references on metal working occupations alone. It is very difficult to evaluate all this work, also the—even more ambitious—applications of vocational testing in pre-war Germany. The authors just mentioned admit that a large proportion of the publications are merely descriptive, containing no convincing validatory evidence, and that their primary object seems to be to "sell" tests to industrialists. In most instances, too, the samples of workers studied are small, the criteria ill-defined and poor in reliability. In the present writers' opinion the results of no single experiment, even on a hundred cases, should be accepted at its face value. To be acceptable, a battery of selection tests should be used over a considerable period and followed up in such a manner as to prove its worth at least twice. For when a number of tests are tried out on a smallish group, some of the coefficients are almost sure to appear promising; yet when repeated on a larger scale, possibly under slightly changed conditions of work or with different types of workers, the results may alter entirely. We certainly do not wish to decry all the work done on specialised tests, even if much of it has little permanent value. The decision between such tests on the one hand and reliance on general tests + interview on the other hand, obviously depends chiefly on whether the investigator is more interested in selection or in classification and guidance. But we can say that the evidence regarding the practicability and validity of specialised tests is hardly convincing enough to dissuade the psychologist from making the best possible use of the simpler techniques, and resorting to the more complex only in so far as they can be proved to add appreciably to the accuracy of his predictions.

The bulk of vocational testing has been in manual occupations, and no clear trends are discernible in other fields. In motor driving, in telegraphy and in music the most successful tests appear to be of the work-sample or miniature situation type. Analytic tests such as Seashore's tests of musical talent have been used, but generally show poor validities. According to Johnson (1946), a simple combination of relevant biographical items is as effective in picking motor drivers prone to accidents as are any of the proposed apparatus tests of co-ordination, reaction time and the like. Useful tests of visual perception for range- and height-finders were developed by American psychologists during the war, but the application of tests of dark adaptation in eliminating men who could not see effectively at night appears to have been a complete failure.

Tests of medical, scientific, legal and nursing aptitudes have usually been validated against examinations in these subjects, not against skill on the job. Thus both in these fields, and in the selection of clerical workers, paper-and-pencil tests either of general or of more specialised abilities are found quite useful. For example, the efficiency of administrative civil servants was successfully predicted by a battery consisting of tests of general intelligence, knowledge of current events, knowledge of the Civil Service, interpretation of charts and tables, and judgment of administrative situations (Mandell and Adkins, 1946). A promising line has been opened up by recent developments in diagnostic testing. We know that different types of mental patient can to some extent be distinguished by projection, sorting, abstraction and other tests; and Thurstone's (1944) factorial analysis of perceptual tests has revealed "qualities of mind" over and above *g*, *v*, and *k*, which appear to discriminate such groups as administrators and student leaders. Munroe (1945) has developed a reasonably objective method of scoring the Rorschach inkblots* to yield a measure of emotional instability which correlates well with failure at a university among students of good ability. The same test is claimed to differentiate mechanics from youths more suited to other types of job (Piotrowski *et al.*, 1944). Thus, there is reason to hope that the future will see the isolation and measurement of some of the special qualities possessed by successful officers and leaders of men, by executives and administrators, by teachers, research workers, salesmen, interviewers, and so forth.

---

* Here the Rorschach test was given to groups, but as nearly as possible in the original form. The further modification where multiple-choice responses are provided, so that scoring is wholly objective, is much less successful (but cf. pp. 256).

At the moment, however, while many tests have been tried out in these occupations, there are none that can be confidently recommended. Indeed, it is fair to conclude that successful guidance or selection of such persons has been based chiefly on the interview and background questionnaire, supplemented by tests of general intelligence, and personality and interest inventories.

Follow-up of Tests in the Forces

During 1942–6 there were seventy-six follow-up investigations in the Navy alone, covering over 31,000 recruits, in the course of which some criterion of proficiency was correlated usually with six or more selection tests, often with other data such as source or mode of entry, age, education and civilian occupation; sometimes also with numerous items such as interests or leadership experience taken from the recruits' biographical questionnaires. The occupations included seamen, gunnery and torpedo rates, radar and asdic operators, stokers, stewards, cooks, photographers, safety equipment ratings, cinema projectionists, supply assistants, writers, R.N.V.R. officer cadets, artificer apprentices, electrical artificers, electrical and radio mechanics, motor, engine-room and ordnance mechanics, wiremen, air fitters and mechanics, telegraphists, signalmen, telegraphist air gunners, Fleet Air Arm pilots and observers, naval instructors and W.R.N.S. personnel selection staff. The groups ranged in size from about 30 to 3,000, but the median size was 300. In all these studies the tests were given, or other data collected, on entry and the recruits' success or failure traced later, usually at the end of training. They do not include validatory trials of new tests.

Soon after the introduction of regular selection procedure (the General Service scheme) into the Army, 2,500 recruits in some twenty representative jobs were followed up, and the test battery was modified in the light of the findings. Numerous subsequent investigations were made of jobs where specific information was needed, e.g. in order to set appropriate test standards. In the main A.T.S. follow-up, some 6,000 auxiliaries were studied in the twenty-seven commonest jobs, the median size of sample being 106. Validation results in the R.A.F. are given in Chapter XVI.

The outstanding facts revealed by this work were the value of a

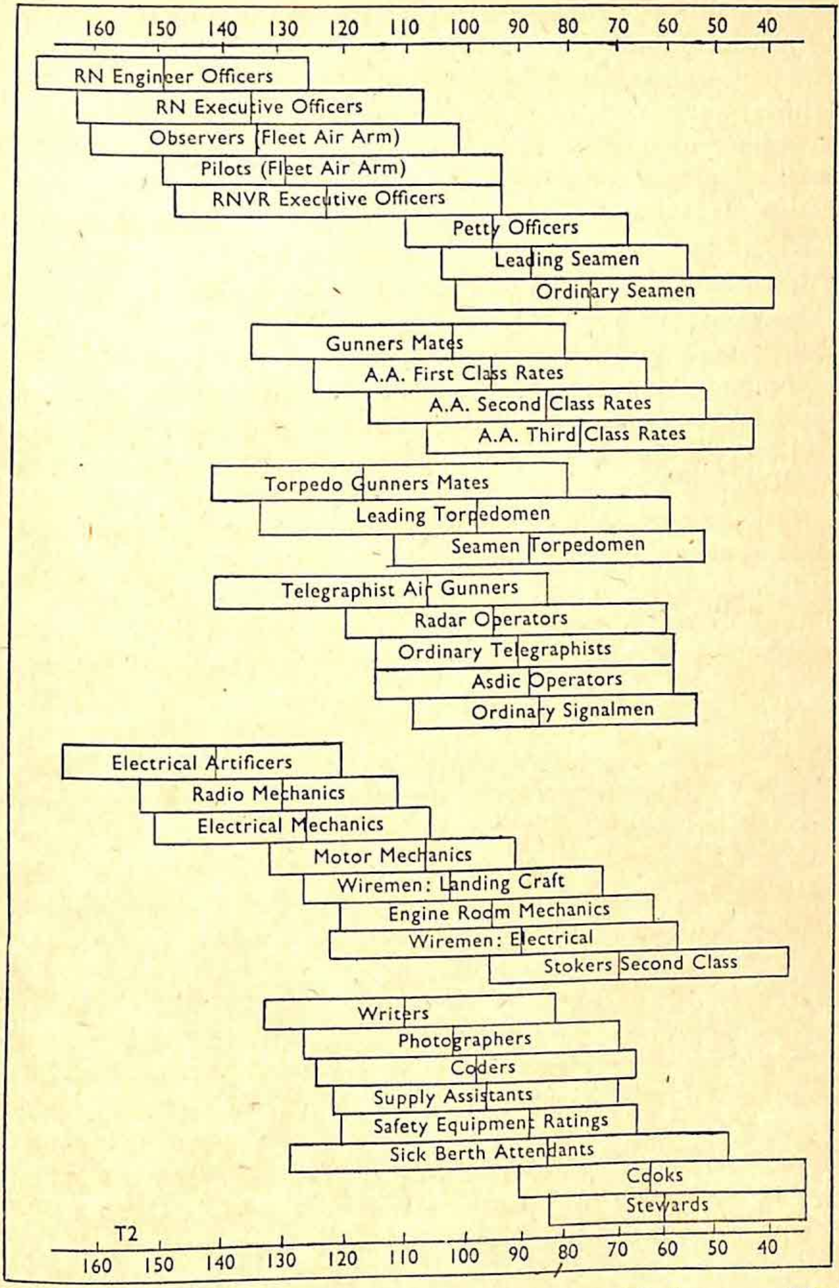


FIG. 4. NINETIETH, FIFTIETH AND TENTH PERCENTILE T2 SCORES OF MEN EMPLOYED IN DIFFERENT BRANCHES OF THE NAVY.

general all-round measure such as T2 or Summed S.G., and the rather small differentiation between different types of job by the more specialised tests. The first point is well illustrated by Figure 4 which indicates the range of T2 among men employed in thirty-six representative naval branches*. The bars show the 90th and 10th percentile scores and the middle lines the medians. Thus the top 10 per cent. of R.N. engineer officers score 168 or over, the bottom 10 per cent. score 125 or under, and the middle man in the group obtained 149. It will be seen that there is excellent differentiation between the more and the less highly skilled branches. For example, less than 10 per cent. of R.N. executive officers and of ordinary seamen overlap, and the same is true of electrical mechanics and stokers, or of writers and stewards.

The results for the main S.P. tests are summarised in the following Tables. Tables XXV and XXVI show the range as well as the median coefficients in all comparable naval and A.T.S. studies. It may be seen that the multiple correlations for the best weighted battery of tests average ·47 in both Services. All the other columns in the A.T.S. Table have been corrected for selectivity, that is, they show the correlations to be expected if *un*selected groups of auxiliaries had been sent forward for training. Thus, the median

TABLE XXV.—RAW VALIDITY COEFFICIENTS OF STANDARD NAVAL SELECTION TESTS

Test	*Matrix*	*Shipley*	*Bennett*	*Maths.*	*Squares*	*T2*	*Multiple r (un-corr.)*
90%ile	·45	·49	·44	·57	·38	·57	·70
50%ile	·28	·30	·28	·35	·22	·40	·47
10%ile	·10	·11	·13	·17	·05	·20	·32

TABLE XXVI.—VALIDITY COEFFICIENTS OF STANDARD A.T.S. SELECTION TESTS CORRECTED FOR MULTIVARIATE SELECTIVITY

Test	*Matrix*	*Bennett*	*Arith.*	*Squares*	*Clerical*	*Multiple r*	
						Un-corr.	Corr.
90%ile	·65	·41	·69	·51	·66	·69	·84
50%ile	·49	·30	·51	·40	·56	·47	·65
10%ile	·27	·19	·26	·20	·35	·35	·50

---

* Almost all of these figures were obtained from groups of recruits who passed their training in 1943. Standards have often changed since then and are very different in the peace-time Navy.

multiple correlation of ·47 rises to ·65 when this selectivity is allowed for (in so far as statistical techniques are capable of making such allowance; cf. Chapter VII). The size of these average or median coefficients is distinctly greater than would have been anticipated from civilian experience, though obviously still too low for accurate selection in the absence of other information. But the Tables show that the correlations varied considerably in different investigations. It was noted that the highest coefficients were usually obtained in jobs involving lengthy training, including a fair amount of theoretical work, where the final assessments of proficiency were based on thorough examinations, and where no scheme of selection run by psychologists was already in operation[*]. The lowest coefficients occurred in jobs where the work is highly specialised (such as radio mechanic), or where previous trade experience is of paramount importance, also in jobs (such as seaman) where assessments of efficiency are based more on personality qualities, e.g. dependability,than on any definite skill or knowledge.

Validities in Different Types of Jobs

We would naturally have expected the verbal and educational tests to show relatively low validities in mechanical and practical occupations, and the mechanical-spatial tests to be of value only among mechanics. But such differentiation was conspicuously small. In some jobs all the tests might achieve high coefficients, in other branches all low, but the *relative* validities of the different tests were remarkably uniform. Tables XXVII and XXVIII compare the mean validities in three main types of work in the Navy and the Army, and while they show that the mechanical and spatial tests were, indeed, relatively less useful than verbal-numerical ones in clerical and communications jobs, yet they were just about as useful among seamen, infantry, officers and gunnery ratings as among mechanics.

In the Navy the Mathematics test obtained the highest validity in most branches and the Squares test the lowest. Often, indeed, the Mathematics test gave better coefficients than T2—the sum of four tests. The Bennett Mechanical test appeared to give useful

---

[*] It is likely that still higher validities would have been obtained had the Forces made any use of objective tests of attainment, as did the U.S. Army and Navy. Ordinary examinations and gradings (Criteria of Types C and D, cf. Chap. VII) are always too unreliable to bring out the full value of the tests. Moreover, the similar make-up of new-type examinations and paper-and-pencil tests undoubtedly boosts the correlations.

TABLE XXVII.—MEAN VALIDITY COEFFICIENTS OF SELECTION TESTS IN CLERICAL, MECHANICAL, AND OTHER NAVAL BRANCHES

Naval Branches	*No. of investigations*	*Matrix*	*Shipley*	*Bennett*	*Arith.*	*Squ'res*	*T2*
Writers, supply, signalmen, telegraphists .	5	·37	·38	·19	·44	·18	·42
Elec. mechanics, air fitters and mechanics, engine room and motor mechanics, artificers, stokers . . .	18	·30	·27	·33	·38	·26	·41
Seamen, officers, gunnery ratings, leading seamen	11	·31	·36	·33	·37	·24	·42

predictions of general practical rather than of specifically mechanical proficiency, and the same was true of the Assembly test in the Army. Although the Bennett test is not usually appropriate for women, yet, together with the "Mec Assembly" test, it showed greater differential value in the A.T.S. than in the men's Services, presumably owing to the lesser diversity among women of previous trade experience. In the Army the Clerical and Verbal tests vied with arithmetic for top place in most branches. Clerical was outstanding, for example, among drivers and in officer selection. In the A.T.S. the same test did best, followed by Arithmetic and spelling. It is noteworthy that American experience seems to have been similar. A large-scale validation of the Army Alpha test in 1918 gave an average correlation of ·54 with officers' rankings of recruits for "value to the Service" (Yoakum and Yerkes, 1920). In a follow-up of some 750 to 1,000 men in six typical naval jobs on

TABLE XXVIII.—MEAN VALIDITY COEFFICIENTS OF SELECTION TESTS IN CLERICAL, MECHANICAL AND OTHER ARMY JOBS

Army job	*No.*	*Matrix*	*Test*						
			2 *Bennett*	3 *Arith.*	4 *Squ.*	17, 25 *Verbal*	12, 21 *Clerical*	8 *Ass.*	16 *Agility*
Clerks, storemen, signallers . .	767	·35	·28	·47	·23	·42	·47	·12	·22
Drivers, linemen, instrument and radio mechanics . .	1,565	·28	·34	·40	·22	·36	·33	·22	—
Gunners, layers, and riflemen . .	1,241	·13	·34	·32	·26	·34	·33	·23	·19

board large American ships, the highest average validity was obtained by an arithmetic test, followed by a verbal intelligence and mechanical knowledge and aptitude group tests. When the U.S. Army revised its general classification test for all recruits, it chose a battery consisting mainly of arithmetic and reading tests, presumably because such tests had proved most valid for Service purposes.

The conclusion followed, in this country, that almost all branches of the Forces wanted the same type of man—one with good education (especially in mathematics) as well as good intelligence. Thus the main use of tests was for apportioning the available supplies of high-quality men among the different branches according to their needs. Differentiation between jobs was based more on interests and interview judgments. Several possible explanations of this result may be proposed:

(1) It is shown below that tests like Clerical and Arithmetic are more reliable than mechanical-spatial or other vocational tests. This alone would help to raise their validities.

(2) Although such tests do depend on education or the $v : ed$ factor, they actually have as high g-saturations as non-verbal tests such as Matrices and will, therefore, correlate well with any job proficiency involving g.

(3) It is possible that success at these tests depends on certain personality or temperamental qualities such as stability and persistence, in addition to g and education, which are relevant to vocational success.

(4) The extreme heterogeneity of recruits in g and education is certainly an important factor. Applicants for any one type of civilian employment would be unlikely to range from university graduates to mental defectives. Most people do not apply for jobs which are quite outwith their capacity. In a more homogeneous group there would be greater scope for specialised tests.

(5) Jobs in the Forces tend to be more variegated than most civilian employments. Both at the semi-skilled level (e.g. seamen, infantry) and at the highly-skilled (e.g. radio mechanics), the recruit is not engaged in one specific type of work. Even when he is mainly concerned with particular equipment and machines he needs a good deal of adaptability to be able to service old, or freshly introduced, models.

Assessments of his proficiency are likely to refer to numerous different duties, hence general tests may be more appropriate than specialised ones.

(6) American Service psychologists have presented evidence from several naval jobs to show that mechanical and other specialised tests are more valid than verbal educational tests when the criteria consist of objective measures of proficiency, but less valid when they consist of instructors' gradings. The implication is that such gradings are based more on written work and on general impressions of brightness than on practical skill. While it is true that the criteria employed in this country consisted largely of subjective gradings (Types C and D rather than Type A), the superiority of verbal tests certainly did not occur only when these gradings involved written work.

(7) Admittedly psychologists in the Forces were able to devote little time to the production of specialised tests and were much handicapped by shortages of material, etc. It is quite possible, therefore, that tests with greater differential value might have been found. Nevertheless, many such tests were tried out—some are described below—but were usually found to add so little to the multiple correlations obtained with the standard tests alone that they were not pursued.

(8) Most of the validatory criteria consisted of training marks or grades rather than assessments of operational efficiency. Thus, tests with educational content might be expected to be the most successful in selecting men who could be trained rapidly. Numerous attempts were made to carry out operational follow-up, none entirely adequate. However, several experiments are worth citing, and so far as they go they do not support this explanation.

Operational Follow-up Results

Assessments of efficiency during fighting in Italy were collected for 200 Royal Marine signallers. Naval selection tests were applied *after* the return of the unit to Britain, and the remarkably high correlation of ·624 was obtained with T2. The numbers of good and less good men in each T2 S.G. are shown in Table XXIX. For the four component tests the coefficients were all between ·46 and ·49. Thus Mathematics appears to retain good predictive value,

TABLE XXIX.—T2 S.G.s OF SIGNALLERS OBTAINING HIGH OR LOW ASSESSMENTS FOR OPERATIONAL EFFICIENCY

Assessment	*T2 Grades*				*Total*
	A	*B*	*C*	*D or E*	
Competent or better	43	66	46	4	159
Passable	1	7	12	2	22
Below standard	1	1	10	7	19
	45	74	68	13	200

but the spatial Squares test does improve when compared with an operational criterion. In Coastal Forces, 185 motor launch and motor gunboat ratings were assessed on several qualities by their officers (by the conference method, cf. p. 110). Correlations with "fighting qualities" were around zero, but with "dependability" the coefficients for Abstraction, Mathematics and Squares were ·25, ·17 and ·30 respectively. Thus in this instance, the spatial test did show up better than *v : ed.* tests.

Some 7,000 men in all Arms of the British Liberation Army were assessed on the questionnaire reproduced on p. 109 at the conclusion of hostilities with Germany, and the test scores which certain groups had obtained on recruitment were traced. Among 260 infantry the Arithmetic test gave a correlation of ·263, and all the other tests smaller coefficients. While this figure is low, the reliability of the assessments may be largely responsible; for the correlation of the operational assessments with efficiency gradings given shortly before the invasion was only ·297.

In the A.T.S. over 600 trainees for anti-aircraft duties were followed up, and later some 1,300 were assessed for efficiency after serving two or more years. The average validity coefficients (corrected for selectivity) are shown in Table XXX. Most of the tests, and the multiple coefficient, drop considerably in validity at the operational stage. But Spelling achieves a much higher

TABLE XXX.—MEAN VALIDITIES OF A.T.S. SELECTION TESTS IN SEVERAL ANTI-AIRCRAFT JOBS AT DIFFERENT STAGES

Test	*Matrix*	*Bennett*	*Arith.*	*Squares*	*Clerical*	*Spelling*	*Multiple r*
Training stage	·47	·24	·53	·11	·42	·06	·65
Operational	·35	·21	·30	·25	·37	·31	·43

coefficient, and the Clerical test is relatively more valid than at the training stage.

Another approach was to contrast the test scores of men promoted to higher naval rates with lower rates, on the assumption that the Navy would select the most efficient rather than the most educable for advancement. The T2 differences among gunnery and torpedo rates are portrayed in Fig. 4, and these, together with the differences on other tests are expressed as correlation ratios (which are similar to correlation coefficients) in Table XXXI. It happens that the absolute correlations are much higher in the torpedo than the gunnery branch, but the relative order of validities is the same in both, and the mathematics test greatly surpasses the others, including T2.

TABLE XXXI.—CORRELATION RATIOS FOR NAVAL SELECTION TESTS

Naval Rates	*No.*	*Shipley*	*Bennett*	*Maths.*	*Squ'res*	*T2*
Four A.A. gunners rates .	1,336	·14	·27	·38	·15	·31
Three torpedo rates. .	264	·31	·45	·58	·39	·53

Conclusions

Most of the above evidence appears to show that the uniformly high validity of verbal-educational tests is not primarily due to the use of training results as validatory criteria, and the explanation must be sought in the other seven reasons. The volume of evidence as to the value of general group tests, and particularly of tests such as Arithmetic and Clerical, is so large and is so unanimous from all the Fighting Services that we can safely say that similar tests would be of great value for vocational procedures in peace-time education and industry, with the following provisos:

(1) The tests must be carefully constructed to suit the populations and the conditions under which they are applied. In succeeding Chapters it is shown that several of the currently available tests were inappropriate, especially among the duller strata of the population.

(2) Such tests are most suitable among school-leavers and other heterogeneous groups. They should not be expected to show the same validity in the selection of applicants for any particular job.

(3) The motivation of the testees must be adequate, and they

should, as far as possible, have had equal amounts of previous practice with tests.

(4) Such group tests chiefly indicate the all-round proficiency, adaptability, and acceptability to employers, and not any particular skills. In other words they show the general level of occupation for which a candidate is fitted, and do not readily differentiate between different occupations at the same level. At the same time it is easier to determine this level and to add supplementary tests where needed, than to follow the principle of preparing an elaborate battery of specialised tests for each specific occupation. When full weight is given to interests, previous experience, etc., as well as to this general level, it may often be found that there is not very much left over which requires testing.

CHAPTER XIII

VERBAL INTELLIGENCE AND EDUCATIONAL TESTS

Abstract.—Ordinary verbal intelligence tests are shown to be less appropriate for adults (particularly those of below average ability) than tests of the abstraction type. An account is given of the merits and defects of clerical tests, oral directions, vocabulary, reading comprehension, verbal fluency, dictation and spelling, arithmetic and mathematics, and dial and instrument reading tests. Examples are given of the tests found most useful in the Services. Selected follow-up investigations illustrate the value of such tests among Army lorry drivers, naval asdic operators and writers, and R.N.V.R. officer cadets.

Most of the intelligence tests in common civilian use consist of batteries of sub-tests, each sub-test containing numerous choice-response items such as vocabulary, analogies, classification, completion, reasoning problems, etc., and having a strict time limit of, say, two to ten minutes. Sometimes, in the so-called omnibus test, items of all types are mixed up, and are explained and practised before the test proper begins.

Several such tests were quite widely applied in the Services, including the N.I.I.P. Group Test 33, and the first parts of the F.H.3 or F.H.R. tests and of Heim's (1947) A.H.4. An omnibus test known as V.I.T. (Verbal Intelligence Test), containing 120 items to be answered in twenty minutes, was devised for the Army. While these were quite useful in testing high-grade personnel such as officer candidates—indeed, V.I.T. was probably the best of the standard W.O.S.B. tests—they were not found satisfactory among average and low-grade recruits. They tend to be unduly wordy and dependent on the testees' literacy, and a considerable proportion of testing time had to be spent on giving instructions and trying out sample questions. This added to the responsibility of the tester, particularly when separate timing of short sub-tests was involved, and as it was necessary to employ large numbers of not very thoroughly trained or experienced testers, the results appeared sometimes to be seriously affected. Another drawback was that the conventional test items and their choice-responses often struck senior officers as artificial and quite unrelated to the soldier's or sailor's jobs. It was very necessary in the early days of personnel selection to win their good will, and it was not easy to answer the criticism which they sometimes raised, namely, that tests of this type are chiefly constructed by university psychologists for testing school-children, and are inappropriate for men and women who have long grown unaccustomed to paper-and-pencil work and to manipulating verbal symbols. Our results on changes in ability between 14–18 years (Chapter XI) lend some support to this view. Factorial analyses showed that V.I.T. and similar tests are good measures of *g*, but that they are also dependent on verbal-educational ability. Thus it is misleading to regard them purely as measures of intelligence. Finally, the fact that they mostly have to be done at maximum speed is objectionable, for, as Cattell (1943) shows, the performance of adults is more seriously affected by age in speeded tasks than in tests with a generous time limit. A novel form of verbal test, known as abstraction, appeared to provide the answer to these difficulties.

Abstraction Tests

Early in the war a pair of short tests of vocabulary and abstraction was published in America by Shipley (1940). They were intended for mental or brain-injury patients whose reasoning capacities often deteriorate more seriously than does their "crystallised" intelligence, shown by knowledge of words. The vocabulary test (S.P.5) was occasionally used in the Forces, but was replaced by the more suitable tests described below. Adaptations of the second test, however, were very widely applied in the Navy (Test 1 and R.C. Test D) and by War Office selection boards (S.P. 45). An abstraction test contains items like the following[*]:

Wherever you see an asterisk, *one* letter or number is missing. Write in the missing letters or numbers.

L M N O P Q *
5 15 25 35 45 **
big little rough smooth hard ****
oz. lb. stone cwt. ***

---

[*] No items quoted here are taken directly from tests which are in use, but are similar to them. The instructions are usually abbreviated.

346 457 568 ***
luck lick lack foul foil ****
quay key owe oh son ***
AzM bXn DvO gTp ***

A remarkable variety of items can be expressed in this form, including the familar analogies, number series, cipher problems, and so on. (Items analyses of several tests showed the alphabetic series and cipher types to be the best of all.) Their content can be given a nautical or military flavour if this is desired. Very easy or very difficult items can readily be constructed, and the instructions needed, even with dull subjects, can safely be cut to a minimum. The reliability coefficients of tests with only twenty items and a ten-minute time limit are quite high in representative adult populations. The scoring is completely objective, i.e. there is no danger of the appearance of alternative or partially right answers, yet at the same time there is none of the artificiality of the choice-response type of test. The test is, indeed, a perfect illustration of Spearman's eduction of relations and correlates. Its dependence on *g* is very high, on verbal ability or education very low. In the writer's opinion this, and the following instructions test, provide the most important developments in group-testing technique made in Britain during the war.

Clerical and Instructions Tests

A clerical test was constructed in 1942 for Army and A.T.S. clerks, which resembled the N.I.I.P. Group Test 25, but which required the testees to perform all the operations—checking, filing, classifying and coding printed information—in rapid rotation instead of in separate sub-tests. This was proved to be a good test for its original purpose, as also was the Institute's test, but to everyone's surprise the new test was found highly effective in selection for numerous other Army categories, as shown in the previous Chapter. There seems to be no necessity to confine this type of test to clerical material, and the following is an example of a possible alternative form:

On each line you have five things to do:

1. Look at the two words at the beginning of the line. If they start with different letters, write a X in column (1). If they start with the same letter, leave a blank.

2. Count up the total number of vowels (A, E, I, O or U) in both words and enter it in column (2).

3, 4. Classify the *longer* of the two words under the following headings:

C. Names of countries.
L. Living things, animals or plants.
N. Non-living objects.

If the first word is longer, write C or L or N in column (3). If the second word is longer, write C, L or N in column (4). Leave the other column blank.

5. In column (5) write a X if the two words belong to the same class. Leave a blank if they are different.

Go through the three examples below, and then work *across* each line*, answering as many of the thirty lines as you can in ten minutes.

		(1)	(2)	(3)	(4)	(5)
France	England .	X	4		C	X
Motorcar	Wales . .	X	5	N		
Elephant	Emerald . .		6	L		
1 Chair	Bed . . .					
2. Rose	Russia . .					
Etc.	. . . .					

It is difficult to conceive why such a test should be successful, for it appears at first sight to suffer from most of the defects of verbal tests listed above. Yet factorial analysis proves that it is as good a test of general intelligence as abstraction and better than Matrices. It is not greatly affected by education, but does involve a purely clerical factor to a small extent. It seems to depend on :

(*a*) Comprehending the rather elaborate instructions. For this reason the test, as used in the Army and A.T.S., was rechristened.

(*b*) Learning the instructions rapidly, in order to avoid going back to the beginning and consulting them frequently.

(*c*) Good vocabulary and ability to abstract the meanings of the words under the given headings.

(*d*) Mental flexibility and speed, or ability to change over quickly from one operation to the next.

---

* It is desirable to make the operations inter-connect as closely as possible, otherwise testees will be tempted to work down each column in turn.

(*e*) Thinking out efficient methods of work, e.g. doing two or three operations simultaneously.

(*f*) A temperamental factor of drive, for maintaining the necessary concentration throughout.

Although (*d*) and (*f*) would not normally be regarded as components of general intelligence, they may enter into many occupations, and so enhance the vocational value of the test*.

Directions Tests

The ordinary written directions item does not appear appropriate for adults of sub-average intelligence, for example:

If the fourth word in this sentence is longer than the fifth word, write down the second letter of the last word; if not, write down the last letter of the third word.

But an oral directions test where complex instructions are read out by the tester was used successfully in the Army Alpha battery and, in this country, gave useful predictions in several jobs such as asdic operator and motor driver. Men who are unaccustomed to reading and writing are much less handicapped than with a printed test. Moreover, efficiency in the Forces depends so largely on understanding of, and prompt reaction to, oral directions that this test might be regarded as a work-sample vocational test. Unfortunately it suffers from one fatal defect, namely, its dependence on the tester's enunciation and manner of application. It can hardly be used unless always given by the same tester, and under really quiet conditions. It was not possible to standardise the test by putting the directions on to a gramophone record, since adequate sound reproduction could not be ensured in all testing centres. Moreover, in one experiment with dictation tests, gramophone records yielded distinctly poorer scores than oral testing (even though the recording was made by a B.B.C. announcer).

Vocabulary and Reading Tests

A test which explicitly involves ability with words was needed for several jobs, including clerks and officers. In the Royal Navy, a reading comprehension test (S.P.96), borrowed from the U.S.

---

* One defect is the abnormality of its score distributions. Nearly 10 per cent. of a representative population can hardly get started on it, and almost as many at the top end get nearly perfect scores, unless the time limit is reduced. Such a U-shaped distribution is advantageous for most selection purposes, but is highly inconvenient to the statistician.

Navy added considerably to the standard tests. This test contains six passages and thirty choice-response questions requiring the abstraction of information from the passages. A much simpler test (S.P.22), in two parallel forms, was devised for semi-illiterate Army recruits and standardised on children to yield reading ages. This was used in an experiment on the effects of basic education courses, where it was shown that the improvement among illiterates sent on a six-weeks' course by their own units was small and irregular; but that among men selected by P.S.O.s and psychiatrists as likely to benefit from the courses, the improvement was much larger. Two tests of attainment in English, chiefly involving reading comprehension and vocabulary, were included in the R.A.F. air crew batteries and one of these (Gen-B) was found to give promising predictions in all categories except pilot.

For individual testing, e.g. of naval neuropsychiatric patients, the vocabulary, comprehension and similarities tests of the Wechsler-Bellevue scale (S.P.6) were used (cf. Trist, 1941). With Army and A.T.S. officer candidates, the Mill Hill vocabulary test (Raven and Walshaw, 1944) was satisfactory. But when older, serving officers came up for re-allocation both this and other verbal tests often seemed to arouse a good deal of anxiety, and a more suitable test was devised consisting of fifteen words (such as FORM and BIT) for each of which four different meanings are to be given in writing in fifteen minutes. This test, the abbreviated Wechsler, and Shipley's vocabulary, all have much the same factor content, showing fairly high dependence on *g*, but also having large verbal-educational loadings.

More reliable than any of these was the Army's Verbal test (S.P.25), based on synonyms, homonyms, and rhymes. Specimen items are shown below:

> Write down on the dotted line a word which means nearly the same as the word on the left, and which starts with the two or three letters that follow:
>
> Example COMMENCE...... BEGIN......
>
> 1. THEFT........... LAR......
>
> Write a third word on the dotted line which rhymes with the left-hand word and means nearly the same as the right-hand word:
>
> Example RANGE...........Change...............ALTER
>
> 1. PIG.................................LARGE

Write a third word on the dotted line which means nearly the same as the left-hand word, and, in a different sense, means nearly the same as the right-hand word:

Example RELISH...........Sauce............IMPUDENCE
1. FIRM......................................RAPID

Like the abstraction test, this has the advantage of objective scoring without choice-responses. (Indeed, Shipley's original test contains an item of the third type just quoted.)

The main R.A.F. battery, devised by Stephenson (cf. Chapter V) has a V or verbal section. This too, contains synonyms and opposites, reading comprehension, and questions on verbs, nouns, and adjectives, e.g.:

Make adjectives (describing words) from the following nouns (naming words). The first has been done for you.

Example LEAF Leafy
1. HEAT ....
2. AFFECTION ....

Dictation and Spelling Tests

Ordinary passages of prose are unsuitable for dictation, since so much time has to be spent on easy, non-discriminating words. Several parallel passages (S.P. 70–74) were devised for the Navy, each containing thirty-five words, almost all of which were fairly to very difficult. But these were barely long enough to be reliable, and, although satisfactory when given by carefully trained testers, were undoubtedly affected to some extent by their enunciations and accents. In one experiment an experienced A.T.S. officer gave the same passage twice, first with good delivery, secondly, with bad delivery which did not involve mispronunciation nor mistiming, but merely poor articulation and failure to say each syllable clearly. The average errors rose from 1·55 to 7·55 under the second condition. Examples of the mistakes that occurred with bad, but not with good, delivery are:

Successive for excessive
In ,, and
Is ,, has
The ,, though
Gratified ,, gratifying
Patrolling ,, patrol.

A better form of test (S.P. 127–130) consists of twenty words, of a suitable range of difficulty, each incorporated in a sentence which helps to define the word and to determine its tense, part of speech, etc. The whole sentence is dictated, but only the critical word is written down and scored. For example:

ONIONS. I grow ONIONS on my allotment.
SEIZE. SIEZE hold of the rope and pull it.

A speed of twenty seconds per word, including its sentence, was found adequate. Even in this test the influence of the tester was apparent. For example, five testers, dictating PLUMBLINE to comparable groups of about 130 recruits, obtained 40, 50, 41, 34 and 51 per cent. of right answers. Presumably the second and fifth testers pronounced the word much more intelligibly than the others, perhaps sounding the B.

Objective spelling tests are the only way out of this difficulty. In the A.T.S. test (S.P.14), five incorrect versions and one correct version of each word were mixed up. A synonym at the beginning of each line helped the testees to identify the word. The right spellings had to be underlined. For example:

RAPID	qick	kwick	qwick	quick	quic	cwik
GRASP	seize	sieze	sease	seez	size	siese

Admiralty psychologists adopted an intermediate form consisting of sentences in each of which one word was mis-spelt, the testee having to find this and write it out correctly (R.C. Test C). For example:

He is a very qwick runner
Sease hold of the rope and pull it.

In one investigation, six varieties of dictation or spelling tests were given to the same recruits, and it was concluded that they all measure the same ability almost equally well. (This conclusion would probably not be true among school-children.) All have moderate *g*, but larger verbal-educational saturations. Straight dictation, however, probably involves least *g*, and the A.T.S. type of test may depend to a slight extent on a distinct factor of ability at clerical work. The straight dictation was the least reliable, but the reliabilities of all other types of test were almost equally high. The recruits' preferences were recorded, and again there were no outstanding differences. In another factorial study the sentence dictation and A.T.S. Spelling tests were shown to measure the same abilities as the Army Verbal test. In a representative male adult population ability to spell does not seem to be differentiated from general education in words.

Arithmetic and Mathematics Tests

Both the Admiralty and War Office psychologists constructed tests in two parts, the first consisting of a few minutes of straightforward addition, subtraction, multiplication and division sums, the second including thirty to forty brief questions for which eight to ten minutes was sufficient time. For example:

How many pence in half-a-crown? ....
$1\frac{1}{3} + 2\frac{1}{6} =$ ....
Increase 50 by 10 per cent. ....
$\frac{\text{sec A}}{\text{cosec A}} =$ ....

Scores on the first part tend to be negatively, on the second part positively, skewed. In other words the lower half of the population is best differentiated by its skill in rote arithmetic, the upper half by its knowledge of fractions, decimals, algebra, etc. The average male adult can scarcely manage simple fractions and decimals, and is completely stumped by percentages, square roots, and the like. The two parts combined provide a highly reliable test which not only spreads out the population better, but is also generally more predictive, than the typical schoolmaster's examination, which is chiefly based on elaborate arithmetical problems, and which is four to ten times as long. We have already commented on the surprising vocational value of these tests and the fact that, even in relation to practical criteria, they often gave better predictions of proficiency in mechanical trades, and among seamen, infantry, and officer cadets, than did either mechanical or general intelligence tests. The second part was most successful among high-grade groups (electrical mechanics, officers, etc.), and the two parts together among average recruits. The high *g*-saturation, particularly of the second part, suggests that more intelligent adults both learn and retain more mathematics than dull ones, and that the test thus measures educ*ability* rather than merely education received. Though both parts do involve a distinct numerical-education capacity, they may also be influenced by temperamental factors. For it has been noted that emotionally unstable and delinquent children at a child guidance clinic tend to be more retarded at arithmetic than at other school subjects. Thus when we find that recruits, picked out by naval instructors as never likely to make good seamen, are poorer on this test than on any other, this does not mean that the instructors cannot tolerate the badly educated, but rather that most of these men are either innately dull, or maladjusted individuals who fail to settle down to naval life, just as they failed to settle to school life. Similarly, Lummis (1946) has shown that bad Army conduct records are associated more closely with low scores on arithmetic than on other S.P. tests, and that arithmetic is most seriously affected by irregular schooling and truancy. It was interesting to observe moderately high correlations even with assessments of officer quality and power of command among R.N.V.R. cadets. Thus when the top quarter of a group of 550 assessed most highly was contrasted with the lowest quarter, it was found that 56 per cent. of the former and 30 per cent. of the latter were acquainted with Pythagoras' theorem, 78 per cent. and 58 per cent. knew the cube root of 64, and so on.

An alternative type of test (S.P.23) was preferred in the A.T.S. because the mathematics test was too difficult. (The average female adult cannot even manage decimals.) Worked sums were provided in which there were three mistakes which the testee had to correct. For example:

	£	s.	d.
Subtract	456	12	8
	269	5	10
	£195	6	8

Such a test, with ten sums to be checked in ten minutes, is also very reliable, and measures nearly the same abilities as the others, but depends to a small extent on ability at clerical work. From the vocational viewpoint, arithmetic was often less predictive than spelling among women.

For its high-grade air crew recruits, the R.A.F. adopted several more difficult American tests, all of selective-response type. Mat-A tests algebra and trigonometry, Mat-B contains verbal problems, and Mat-D rote arithmetic. Mat-C requires testees to find approximate answers to such questions as:

> 248·4 miles per hour for 20 hours and 15 minutes = (*a*) 513 miles, (*b*) 50 miles, (*c*) 5,000 miles, (*d*) 500 miles, (*e*) 503 miles.

Mat-F involves reading off the correct entries from numerical tables. The two latter tests bear a close resemblance to some of the jobs of the navigator.

Instrument Reading Tests

Two other tests borrowed from the U.S. Army Air Force were found useful in the allocation of air crew categories. Ins-A shows sets of dials, as on an aeroplane panel, marked Altitude, Fuel-Air Ratio, etc. There are fifty-seven choice-response items calling for the correct reading of appropriate dials. In Part I of Ins-B sets of six dials have to be read and related jointly to descriptions of an aeroplane's velocity, orientation, altitude, etc. Part II shows sets of two dials—Artificial Horizon and Compass. These have to be read and related to illustrations of aeroplanes in flight. In each item there are five planes, only one of which can be reconciled with the pair of readings.

A simpler test (S.P.119) was devised for naval radar operators. The first part shows several pictures of bits of rulers, scales and dials, the testee having to read off the numbers indicated by pointers (with or without interpolation). The second part involves reading the co-ordinates of points on a rectangular, and on a very simple polar, graph.

Illustrative Investigations

Many of the above statements and conclusions about tests may appear somewhat dogmatic. It is impossible to give all the evidence upon which they are based, but the following brief outline of some of the main validatory experiments will, it is hoped, cover some of the ground.

R.A.S.C. Lorry Drivers.—A large battery of tests gave the following correlations with driving proficiency in a group of 240 men followed up at the end of training.

TABLE XXXII.—VALIDITY COEFFICIENTS OF TESTS FOR DRIVERS

Age	·250
Matrices	·312
Bennett Test 2	·586
Personality inventory	·439
Previous driving experience	·511
Cycling experience	·382
Oral directions test	·486
Questionnaire on driving knowledge	·431
Interest blank scored for drivers	·423
Group choice reaction time	·335
Squares, Test 4	·287
Interest blank scored for mechanics	·266
Judgment of distances	·212
Judgment of ellipses	·200
Judgment of speed	·150

In another similar group of 306 recruits Chleusebairgue's (1939) elaborate choice reaction time test gave a correlation of ·385. The multiple correlation of the first four items on the list combined is ·669, and the figure does not rise to more than ·68 or ·70 when any other test (except driving experience) is added. Note that the three analytic tests, borrowed from the N.I.I.P. battery (Miles and Vincent, 1934) come at the bottom, and that oral directions is the best test after Bennett. The interest blanks and inventories are discussed in Chapter XV.

Asdic Operators.—Several experiments were carried out in anti-submarine training schools, on the basis of which an omnibus intelligence test (Selection Test A), mainly consisting of oral directions items, a gramophone test of sense of pitch (Doppler effect), and a group choice reaction time test, were installed as a regular battery. It was also found that Heim's A.H.4 test, containing verbal and non-verbal intelligence items, was a satisfactory alternative to Selection Test A. Later an individual audiometric examination was added by the medical officers. One sample of 282 trainees, selected in respect of the above battery and A.H.4, but not directly selected on T2, yielded the correlations with training results shown in Table XXXIII.

TABLE XXXIII.—THE VALUE OF DIFFERENT INTELLIGENCE TESTS IN SELECTING ASDIC OPERATORS

	Correlations	*Corrected for Selectivity*
Asdic battery	·343	·553
A.H.4	·384	·572
T2	·440	·597

When statistical correction for selectivity was applied, all three tests or sets of tests showed similar, fairly high, validities.

TABLE XXXIV.—THE VALUE OF INTELLIGENCE AND OTHER TESTS IN SELECTING ASDIC OPERATORS

Tests	*Correlations*
Intelligence, Selection Test A	·417
Group choice reaction time	·045
Doppler Record	·179
Audiometric threshold	·145
Audiometric sense of pitch	·165

In the course of time the reaction time apparatus, and its method of application, became unreliable. Nearly four years after the institution of the tests the correlations with training results, shown in Table XXXIV, were obtained among 358 recruits (also a selected group). It will be seen that the intelligence test retains much the same validity as before, and the Doppler record assists slightly in selection. The audiometric tests, though also useful, do not appear superior to this group auditory test.

Writers and Supply Assistants.—Some 800 of these naval clerks were followed up in 1943, and their scores on the standard battery correlated with training results. Four to six hundred more were given additional tests in 1944–5, with the results shown in Table XXXV. (Tests 2 and 4 are omitted since they gave much lower coefficients.) The second set of figures for T2 tests is lower than the first because of more efficient selection, and the additional tests attain rather higher coefficients than they would have, had they also been used for selection.

TABLE XXXV.—THE VALUE OF TESTS IN SELECTION FOR CLERICAL WORK

Test	*Correlations in:*	
	1943	1944-5
0 Matrix	·37	·27
1 Abstraction	·36	·29
3a Arithmetic	·36	·29
3b Mathematics	·43	·32
T2	·42	·36
71 Dictation	·36	·33
21 Army Instructions	—	·43
Group Test 25 (N.I.I.P. Clerical)	—	·44
96 U.S. Reading comprehension	—	·38

Nevertheless, it was found that either of the clerical tests, or the reading comprehension, would add appreciably to predictions based only on the standard battery.

R.N.V.R. Officer Cadets.—T2 was regularly applied to officer cadets under training in H.M.S. *King Alfred* in 1942–3, but several additional tests were tried out on successive large groups, mostly around 500, which were not taken into account by the Admiralty selection boards. The additional variance in final passing out marks covered by these extra tests was calculated, but, for simplicity's sake, the results are expressed in Table XXXVI in the form of partial correlations with T2 (in the first column) or Tests 1 + 3b (in the second column) held constant. The validity of T2 itself ranged around ·535, and the validities of the additional tests were between ·30 and ·55.

TABLE XXXVI.—PARTIAL CORRELATIONS SHOWING THE ADDITIONAL VALUE OF VARIOUS TESTS IN OFFICER SELECTION

Test	*Partial Correlations:*	
	T2 Constant	1 + 3*b Constant*
96 Reading comprehension	·20	
Mathematics examination	·17	
12 Army Clerical	·21	
21 Army Instructions	·27	
Group Test 25, N.I.I.P. Clerical	·17	
45 W.O.S.B. Abstraction	·10	
15 V.I.T.		—·04
Group Test 33, N.I.I.P. intelligence	·19	·13
5 Shipley Vocabulary	·13	
109 Figure construction		·11
97 Memory for designs	·05	
80 Orientation	·25	·24
Group Test 70/23 non-verbal	·03	
Progressive Matrices, 1942	·06	

The Clerical or Instructions tests certainly have a contribution to make, and are probably superior to the conventional form of clerical test (Group Test 25). Reading comprehension, too, is useful, and possibly the mathematics examination set by the training establishment itself. Other verbal intelligence tests are of more doubtful value, and non-verbal ones appear quite useless. Two spatial tests add very little, but a verbal orientation test (discussed in the next Chapter) seems to be one of the best. Unfortunately, it was given to only 155 cases, hence its results are less certain.

CHAPTER XIV

NON-VERBAL AND MECHANICAL TESTS

Abstract.—The Progressive Matrices and other non-verbal intelligence (*g*) tests were widely used in the Services. Several tests of spatial judgment (*k*-factor) were tried. Though these are on the whole less valuable among adults than adolescents, e.g., engineering apprentices, there is some evidence that they are predictive of "practical" ability. Performance tests were developed for application to African and Indian recruits and other special cases. Orientation and observation tests are described. Among the various mechanical tests, group paper-and-pencil tests of mechanical comprehension, mechanical and electrical information and trade knowledge, were generally more predictive than practical assembly or trade tests. Illustrative investigations show some of the results obtained among naval radar operators, boy tradesmen, radio and electrical mechanics.

The Progressive Matrices test (Raven, 1939) was adopted as the primary general intelligence test in the Royal Navy, Army and A.T.S. in 1941, largely in order to ward off criticisms of the educational bias in verbal tests. An item of the Matrices type is shown in Fig. 5. This test has been applied to greater numbers of men and women in this country than any other single one. For ease of administration a twenty-minute time limit was usually imposed. A harder version (unpublished) was constructed by Raven for testing Army officer candidates in 1942. Numerous factorial analyses showed that, while Progressive Matrices is an almost pure *g* test, it does involve the visuo-spatial or *k* factor to a small extent. For vocational purposes it was somewhat disappointing. Three reasons may be suggested. First, the success of such tests as instructions and Arithmetic showed that the Services did not actually want pure intelligence so much as intelligence and education. The comparatively small extent to which it differentiates between different occupational grades has been pointed out in Chapter XI. Secondly, whenever a battery of verbal and mechanical-spatial tests was

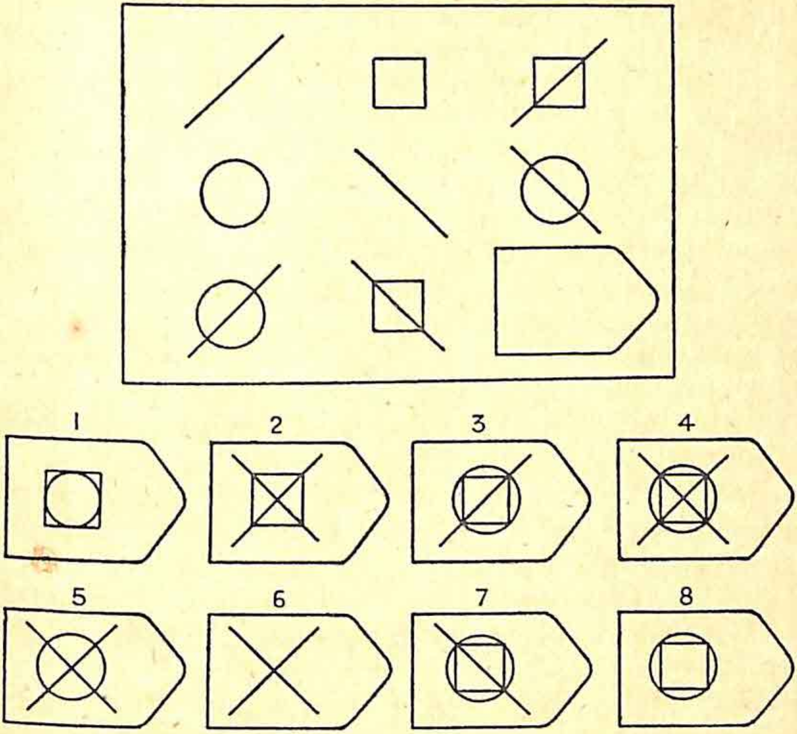


Fig. 5. Specimen Item as in the Progressive Matrices Test. Which of the Numbered Pieces 1-8, if Fitted into the Empty Space Above, Would Complete the Pattern?

applied, their combined result yielded a considerably better measure than did Matrices alone. Only when its visuo-spatial component was involved did it seem to have anything to add to such a battery. Thus its best correlations were with proficiency in visual signalling and radar operating and among A.T.S. recruits engaged on various mechanical and anti-aircraft duties. The third reason is its rather poor reliability, and its susceptibility to non-intellectual influences. More than any other group test it is affected by age and other types of mental deterioration, possibly also by temporary emotional stress. The improvement in scores after attendance at physical development courses was described above (p. 201). Dr. J. A. Fraser Roberts was able to show, by a detailed comparison with more reliable tests, that the unreliability is greatest in the 15–30 score range, i.e. just about the level where acceptance or rejection for the Services takes place. Its efficiency is greater the higher the score (in spite of the negative skewness of its distribution), and also at very low score levels. Doubtless, the test is more valuable when given without time limit by an experienced clinical psychologist who can interpret the significance of failures on particular test items. No evidence could be obtained, however, that men with scores which Raven calls "unreliable", i.e. irregular patterns of scores on the five sections of the test, were more neurotic or in any other way different from men with "reliable" scores (cf. Eysenck, 1944). The test is seldom given now in the Services, mainly because its wide use has caused it to become rather well known.

Another matrices test, where the testee draws in his answers instead of choosing one of a number of patterns, was devised by Stephenson for the *g* section of his R.A.F. battery. Though sometimes difficult to score, this version seems to be more reliable. The N.I.I.P. Group Test 70, which was occasionally used in the Navy and Army, has three parts of which the last two resemble Progressive Matrices. Part I is a very simple but surprisingly effective test. Five two-line drawings are shown, with a long series of incomplete bits of these drawings. The testee has to identify the complete drawing to which each bit belongs. In the absence of definite evidence, it is possible that this test involves *g* for much the same reason as the instructions test, namely, rapid learning of the complete drawings, mental flexibility, and continuous concentration. Like Matrices, all parts show some dependence on visuo-spatial ability. An alternative test to Matrices was constructed for the Army from domino patterns. This Dominoes test, though less attractive to the testees, is more reliable and more *g*-saturated than Matrices and shows no visuo-spatial element, but may involve number-ability to a small extent among dull and backward recruits.

Spatial Judgment Tests

The spatial test most widely used in the Navy, Army and A.T.S. was the N.I.I.P. Squares test (cf. Fig. 6). This consists of a series of fifty figures in each of which the testee has to draw a dividing line such that the two pieces so formed would, if turned around, make a square. The N.I.I.P. Group Test 80 (devised by Slater, 1940) was occasionally given to Army engineer officer candidates and R.A.F. apprentices. This is based on the recognition of shapes when turned around or shown mirror-wise. Though the Institute's

Form Relations test was seldom employed because of its heavy paper consumption the accompanying Memory for Designs (S.P.97) was often given in the Navy. It has the drawback of being very difficult to score. A test based on the copying of straight-line figures on to a space marked out with dots has been used by Stephenson, McQuarrie, Slater ("Figure Construction Test") and others (cf. Fig. 7). An extension of this was to have some of the figures reversed mirror-wise or turned through 90 degrees before being copied. Since the testees' responses are creative instead of selective, there is a slight subjective element in scoring their correctness. Still another type of spatial test is the paper formboard (cf. Paterson and Elliot, 1930). The selective-response versions of this test, sometimes used in America, appear rather artificial. But a creative-response version tried out in the Navy had to be discarded because of the subjectivity of scoring, in spite of the great

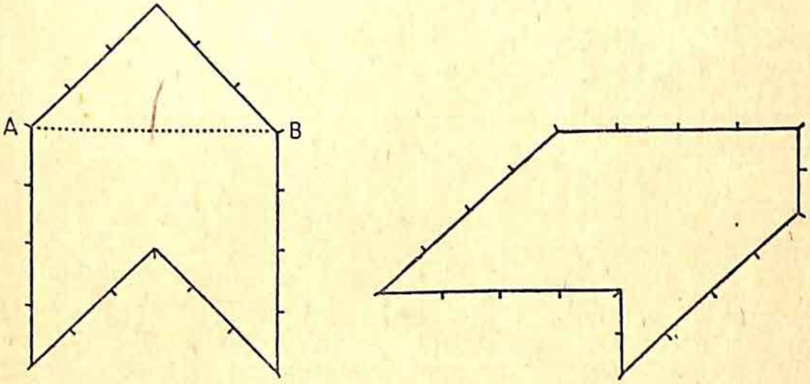


FIG. 6. SQUARES TEST. IN THE LEFT-HAND FIGURE, A LINE DRAWN FROM A TO B WOULD DIVIDE THE FIGURE INTO TWO PIECES WHICH COULD BE PUT TOGETHER TO MAKE A SQUARE. IN THE RIGHT-HAND FIGURE A SINGLE LINE CAN BE DRAWN BETWEEN TWO POINTS, DIVIDING IT INTO TWO PIECES WHICH WOULD MAKE A SQUARE.

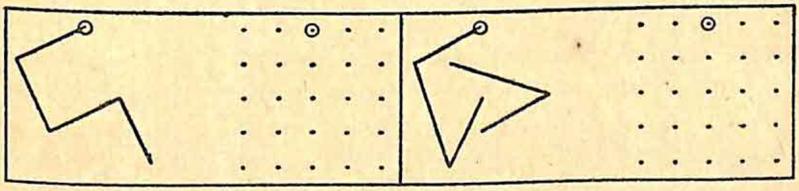


FIG. 7. FIGURE CONSTRUCTION TEST. COPY EACH FIGURE ON TO THE DOTTED SPACE TO THE RIGHT OF IT. EACH LINE MUST START AND END UP AT A DOT. BEGIN AT THE DOT WITH A CIRCLE ROUND IT.

advantage that this test can be "got across" to dull recruits more readily than any of the others discussed here (cf. Fig. 8). Stephenson's *k*-test contains paper formboard, figure construction and mirror-reflection items.

The precise significance of these tests, and the extent to which they measure the same abilities or factors as non-verbal *g* tests like Matrices, or performance and mechanical tests, are obscure. While they were originally designed as intelligence tests (for example in Army Beta), many investigations such as Kelley's (1928), El Koussy's (1935) and Thurstone's (1938) have shown that they embody a space factor—named "*k*" or "*S*"—distinct from *g*, whose essence (according to El Koussy) is the use of visual imagery for the mental manipulation of spatial relations. Investigations in the Services similarly showed that there is little to choose between the half-dozen tests just mentioned. They all depend to about 30 per cent. on *g* and to the same extent on *k*, the remaining 40 per cent. being specific factors or unreliability (none of them is as reliable as a verbal or numerical test of the same length). Price's (1940) study shows that *k* is identical with the practical factor that enters into many performance tests, which Alexander (1935) calls *F*. Though Drew (1947) has recently denied this, his figures actually appear to support Price's. These, and Alexander's own investigations, further suggest that allegedly pure *g* tests may contain some *k* (as we found with Matrices), and that there is considerable overlapping with the "*m*" factor present in mechanical tests.

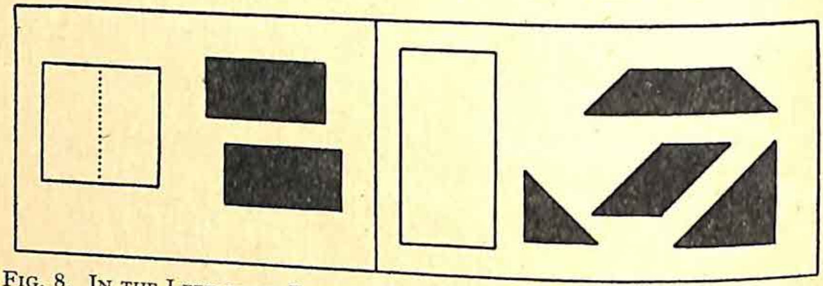


FIG. 8. IN THE LEFT-HAND FIGURE, THE TWO BLANK PIECES COULD BE FITTED TOGETHER AND TURNED AROUND TO GIVE THE SQUARE. DRAW LINES IN THE RIGHT-HAND SHAPE TO SHOW HOW THE FOUR BLACK PIECES COULD BE FITTED TOGETHER.

There is ample evidence (cf. p. 207) of the vocational value of *k* tests among adolescents with little mechanical experience, but very rarely were they found to assist in the selection of adult mechanics in the Forces. Nor did they show any marked correlation with radar operating or other jobs involving prominent visual components. Nevertheless, the occasional successes of Squares were particularly interesting. Not only did it show up better when compared with operational, as distinct from training, criteria among Royal Marine signallers and Coastal Forces ratings, but also higher rates in the regular (continuous service) Navy did unusually well, and advancement correlated with educational ability chiefly among conscripts (hostilities only recruits). In the A.T.S., Squares was outstanding among draughtswomen and certain anti-aircraft personnel. There was fair agreement with practical efficiency tests in infantry. Finally, Army recruits picked out by psychiatrists as lacking in combatancy or questionable in emotional stability did particularly badly in the test. This evidence suggests that there is, indeed, a rather vague and ill-defined factor of practical ability, distinct from *g* and from mechanical ability, which is measured—unfortunately not at all efficiently—by visuo-spatial tests. Shortage of psychologists and restrictions on paper prevented fuller research which might have led to the development of better tests of this type.

Performance Tests

The Cube Construction test was sometimes applied by Admiralty psychologists to doubtful candidates for mechanic branches, and several performance tests were tried out in W.O.S.B.s (cf. p. 56). But all these were regarded more as qualitative than as quantitative tests, since they threw light on the candidates' methods of tackling problems, and no systematic results were collected on large numbers. The performance test most widely used in the Navy and Army was an adaptation by Trist and Misselbrook of Kohs Block Design, in several ways superior to the Kohs (1923), Alexander (1935) and Drever-Collins (1936) versions. A new series of designs was prepared, and scoring was based on the numbers of blocks correctly placed within quite short time limits. This seldom takes more than twelve minutes to give and is highly discriminative from about 9–10 year up to superior (though not very superior) adult level.

Although Spearman and his followers (El Koussy, 1935; Cattell, 1943) regarded performance tests merely as rather inefficient measures of *g*, containing no separate practical factor, Alexander's and other experiments cited above show that this is not true of adolescents and adults. The only analysis carried out in the Forces indicated that Kohs-Misselbrook Blocks measures much the same abilities as Squares, that is $g + k$ or F.

Numerous performance tests were devised for both African and Indian personnel. Though adequate validatory evidence was difficult to obtain, it appeared that the most successful tests were not of the conventional formboard jigsaw or picture completion types (even though the pictures were of objects presumed familiar to the recruits) but adaptations of tests which worked well in this country, namely:

(*a*) Continuation of patterns or series made of strips of coloured wood (analogous to the Abstraction test).

(*b*) Reproduction of figures by sticking pegs into holes (closely parallel to Figure Construction).

(*c*) Simplified Kohs Blocks.

Orientation Tests

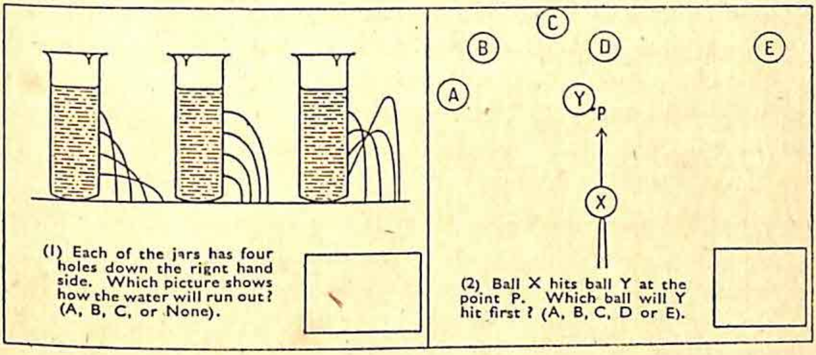
A very different type of spatial test contains verbal problems such as:

> If when standing on your head you are facing East, in what direction is your right arm pointing—N., S., E. or W. ?

Very promising results were obtained with these tests both among Army trainees assessed for orientation or sense of direction, and in predicting the navigation marks of officer cadets in the Royal Navy. They were never developed further partly because the inclusion of sufficient items for good reliability would make them very long, and partly because it is difficult to prevent testees drawing plans unless (as in the Terman-Merrill scale) they are tested individually. Factorial analysis of one such test indicated a high g-saturation, and a spatial component which only overlapped slightly with that of ordinary k-tests.

Observation Tests

Three tests borrowed from the U.S. Army Air Force were used in the selection of air crew. Obs-A shows eight large aerial photographs, on which various objects or small areas of the terrain are lettered. Below each are six small photographs, as it were cut out of the large one, and the testee has to match these with the appropriate lettered areas. In Obs-B small photographs have to be matched with appropriate sections of large coloured maps, which are sub-divided into a dozen lettered districts. Obs-C is an aircraft identification test. Each item shows a silhouette of an aeroplane in the left-hand margin and five other similar silhouettes with the aircraft pointing in various directions. One out of the five has to be chosen which is identical (apart from direction) with the initial silhouette.



(1) Each of the jars has four holes down the right hand side. Which picture shows how the water will run out? (A, B, C, or None).

(2) Ball X hits ball Y at the point P. Which ball will Y hit first? (A, B, C, D or E).

FIG. 9. MECHANICAL COMPREHENSION TEST ITEMS.

These tests provided a good illustration of the dangers of judging from face validity. A and B in particular look as though they would be valid. But, though they did correlate significantly with initial and basic training marks among navigators, the coefficients were disappointingly small. In a factorial analysis all three tests were found to embody g + a group factor, distinct from k, and this might, perhaps, be identified with the P or perceptual ability factor claimed by some American psychologists (Dvorak, 1947). More probably, however, this component is a result of training, some applicants for air crew having had more experience than others in reading maps and photographs and in aircraft identification.

Mechanical Tests

The tests chiefly used in the Forces were:

*S.P.*2, *Bennett Mechanical Comprehension* and revisions thereof. Some forty to fifty pictures are shown of mechanisms (e.g. gear wheels), or mechanical situations in everyday life (e.g. trains going round bends, men lifting weights, etc.), and selective-response questions are asked, as in Fig. 9.

A new *Practical Problems* test for the A.T.S. is based on pictures and questions about cooking, clothes, motor-cars and other things of which women might be expected to have had experience.

Mec-B, *Mechanical Comprehension* (*Diagrams*).—Eight diagrams of machines are shown, and several selective-response questions are asked about the workings of each. This test has generally yielded better validities in the R.A.F. among pilots, navigators and flight engineers than has the pictorial Bennett test, Mec-A.

*S.P.*117*E. and M., Electrical and Mechanical Information.*—Effective ten-minute tests were constructed containing twenty-five electrical and thirty mechanical items of the following, simple-completion, variety:

The terminals in a metal electric lamp holder are usually mounted in.................

To get a keen edge on a chisel the final sharpening is done with.................

It is seldom possible to devise items where only one response is correct (these two were rejected because they produced so many alternatives). Hence scorers have to be provided with comprehensive lists of permissible answers. Nevertheless, they have decided advantages over selective-response tests. The R.A.F. has used both these and selective-response tests of mechanical information (Mec-C) and aviation + engineering information (Gen-D). The latter obtained some of the highest coefficients of any test among pilots and air gunners, and was also the best single test in the U.S.A.A.F.

Correlations between comprehension and information tests are so high (approximately ·7 in unselected groups) that they are clearly measuring much the same ability. Thus items of both types were included in the new naval recruiting centre tests. Questions involving specific trade experience were avoided in Test 117, and a separate series of *Tests of Trade Knowledge* was constructed, including verbal and pictorial questions which had been proved to differentiate between groups of men known to be skilled in that trade and an inexperienced group (cf. pp. 31–2, 45).

*S.P.*8, *Assembly Test.*—A test of the Stenquist type, containing nine sets of parts to be assembled, was devised for the Army. Only parts which could readily be duplicated from Army stores were included. It was so arranged that one tester could test eight recruits simultaneously in about thirty-five minutes.

The A.T.S. required an easier mechanical test and one was devised which involved stripping and reassembling seven Meccano models with the aid of full-sized pictures (S.P.24). In the Navy, a test of ability to bend a piece of wire with pliers into a given shape was used in several investigations (S.P.103), but it was not applied regularly because the scoring against a "quality scale" was somewhat difficult. These tests, too, could be given to small groups. Cox's or Vincent's models tests of mechanical comprehension were applied in certain naval and Army investigations, and in the R.A.F.

All the above mechanical tests and some of the spatial ones were shown to be of value in the Forces in the selection of apprentice tradesmen at 14–15 years, and to add considerably to predictions based on academic examinations and intelligence tests, thus confirming the results of Hunt and Smith (1945) and Holliday (1943). We have already described their disappointing validity among male adults and attributed it, partly to poor reliability, and partly to the distorting effects of the varied mechanical experience possessed by recruits from non-mechanical civilian occupations as well as by tradesmen. Models tests appeared to be even less successful than paper-and-pencil or assembly ones. Since the specifically mechanical content of the latter was of little use, they acted as tests of all-round practical ability in much the same way as —or better than—spatial tests. Thus the assembly test obtained its highest validity coefficients (averaging ·36 and ranging up to ·58) when compared with tests of elementary training among infantry, assault course marks in R.A.C., and proficiency in demolitions, field work, bridging, etc., in the R.E.

Given more paper and materials, and more staff, better tests might perhaps have been developed. An important principle established in later investigations was to choose test items which contributed most to mechanic selection over and above *g* and education, in order to cover different ground from that predicted, say, by Arithmetic and Instructions. But mechanical ability is itself so complex that the search for a test which might be useful for professional engineers, for radio technicians, for fitters, for garage hands and for machine operators, does not seem very hopeful. Thus Admiralty and War Office psychologists came to rely more and more on straightforward information tests. Tests like S.P.117E. and M. were often found to be more predictive than either Bennett or Arithmetic. Tests of trade knowledge were particularly useful in gauging the extensiveness of experience claimed by men who had already entered a civilian trade, and its relevance to Service jobs. While it is true that these are paper-and-pencil tests, they are much more reliable than the conventional practical trade test, and it was sometimes possible to show that they give better forecasts of success in passing mechanic or electrician training courses. Objections were sometimes raised that mechanics selected in this way might be lacking in the "practical knack," but it is extremely dubious how far such an ability really exists. The evidence amassed by psychologists appears to show, rather, that men may possess widely different degrees of practical ability in different jobs. Hence the best basis for prediction consists of records of work actually done. This problem of the nature, and consistency, of mechanical ability will be discussed more fully elsewhere.

Illustrative Experiments

The following validatory experiments illustrate some of the points made above.

Radar Operators.—In 1943 correlations were obtained between the standard naval tests and success at the training course among 603 radar operators. They are shown in the first column of Table XXXVII, corrected for selectivity. This is the only naval study in which the coefficient for Matrices surpassed that for T2, though it equalled it among visual signallers. It was shown, too, that Matrices was most useful at the bottom of the scale for differentiating fails from passes, whereas Mathematics was more useful at the upper end. Failing the course depends chiefly on practical operating whereas high marks are determined to a greater extent by ability at theory. Note that Squares is the poorest test. Apparently the perceptual ability required among operators is not the same as *k*.

TABLE XXXVII.—THE VALUE OF TESTS IN SELECTING RADAR OPERATORS

Test	*Correlations in:* 1943	1945
0 Matrices	·42	·30
1 Abstraction	·34	·27
2 Bennett	·28	·22
3a Arithmetic	·25	·29
3b Mathematics	·39	·35
4 Squares	·20	·14
T2	·39	·36
Group Test 70 (all parts)		·32
119 Scale and graph reading . . .		·42
97 Memory for designs		·24
109 Figure construction		·27
118 Oscilloscope reading		·21

By 1945 the radar sets and the nature of the training had greatly altered, and in a fresh experiment the standard battery along with several new tests was given to 411 radar plot ratings. The (uncorrected) correlations with practical marks are shown in the second column of Table XXXVII. Although Mathematics is now relatively more important, and ability to read scales and graphs gives much the best predictions, Matrices still does quite well. Group Test 70 is similar. Two additional *k*-tests are better than Squares, but do not add appreciably, and a U.S. Navy oscilloscope reading test, based on discrimination of pictures of radar patterns, is of little value.

Army Apprentice Tradesmen.—Approximately 1,000 boys who took examinations and certain batteries of S.P. tests on entry at 14 years were followed up over periods of one to three years, during which they had been trained as fitters (general, motor vehicle, or gun), armourers, electricians, instrument mechanics, carpenters, masons, or in other smaller trades. The courses were sufficiently similar, and the correlations in different trades sufficiently uniform to justify the quotation, in Table XXXVIII, of average figures.

TABLE XXXVIII.—THE VALUE OF TESTS AND EXAMINATIONS IN SELECTING APPRENTICE TRADESMEN

Examination or Test	*Correlation*	*Mechanical Test*	*Correlation*
Arithmetic entry exam.	·15	2 Bennett	·29
English entry exam.	·08	4 Squares	·29
S.P.3A Arithmetic	·13	8 Assembly	·27
S.P.3A Mathematics	·20	(Meccano Assembly	·22)
S.P.17 or 25 Verbal	·16	(Mechanical interests	·37)
(S.P.21 Instructions	·32)		
Matrices	·15		
Oral directions	·13		

While all coefficients are small, this was due largely to the unreliability of the available criteria. When divided into more and less reliable sets, the multiple correlations with the criteria were ·48 and ·30 respectively. The tests listed in parentheses were applied to some 300 boys only, for whom the criteria were exceptionally good, hence their coefficients are not comparable with the rest. Bennett, Squares and Assembly are the most generally useful tests, along with Mechanical Interests—a test on the same lines as Strong's Vocational Interest Blank. Arithmetic is of some value, but general intelligence, Directions, and verbal-educational tests (with the possible exception of Instructions) have little to contribute. The correlations for examinations are unduly low since these examinations were actually used for selection, but correction raises them only slightly.

A similar study with 850 naval artificer apprentices showed that the following tests have the best validities:

S.P.2 Bennett, 117E. and M. information.

Medium validity coefficients were obtained by:

S.P.97 Memory for Designs, 4 Squares, 3b Mathematics, 7 Kohs Blocks, 110 Cox models and 103 Wirebending.

Very low validities were obtained by:

S.P.1 Abstraction, 3a Arithmetic and the entry examination.

Radio and Electrical Mechanics and Wiremen (L).—Several information tests were given to some 300 naval radio mechanic trainees, 214 electrical mechanics and about 100 wiremen (L)—that is civilian electricians who had passed a practical trade test on entry into the Navy. Correlations with training results are shown in Table XXXIX. Test 84 is a test of trade electrical knowledge, 99 a test of knowledge of radio symbols, and 117E. of everyday or amateur electrical information.

TABLE XXXIX.—THE VALUE OF TESTS IN SELECTING ELECTRICAL AND RADIO TRAINEES

Test	*Correlation among:*		
	Radio Mechanics	*Electrical Mechanics*	*Wiremen (L)*
T2	·29	·51	·41
2 Bennett	·32	·24	·29
3b Mathematics	·24	·54	·40
84 Electrical trade knowledge . .	·33	·33	·46
99 Radio symbols	·35	·29	·42
117E Electrical information . . .	·46	·46	·36
117M Mechanical information . .	·21	·30	·30
Initial trade test	—	—	·36
P.S.O.s' judgments of suitability . .	·33	·29	—

All three information tests surpass the standard battery among radio mechanics. With electrical mechanics, mathematics is especially important, but radio knowledge is unnecessary. Thus Tests 3b, 117E, 84 and 117M have the highest coefficients. Among wiremen the trade test is less predictive than any of the information tests, Mathematics or T2. Similarly unsatisfactory validities for trade tests were found in other investigations of electrical artificers (direct entry) and cinema projectionists.

CHAPTER XV

SPECIAL APTITUDE, AND TEMPERAMENT, TESTS

Abstract.—Descriptions are given of tests of physical agility, and of auditory capacities including aptitude for morse; also of tests involving apparatus for gunners, radar operators, aircraft pilots and other special groups. The results obtained, particularly with the simpler tests, were mostly disappointing, and it appeared that more successful selection could be achieved with the aid of paper-and-pencil tests and "work-sample" methods. Some progress was made with the selection of recruits suitable as instructors, but very little in picking those with good or poor personality and temperamental qualities, such as leaders on the one hand, and emotionally unstable or psychiatric suspects on the other hand. Recent developments in personality testing are reviewed. Projection tests and other indirect methods are valuable in skilled hands, and batteries of objective tests for certain "dimensions" of personality such as general neuroticism are feasible, though too elaborate for large-scale use. Neurotic inventories or questionnaires gave good results in the Forces, but would not necessarily work so well under civilian conditions. However, these, together with group projection tests, and interest questionnaires, deserve further investigation.

A number of specialised tests were tried out in the Forces and a few were used regularly, though none was outstandingly successful. Some of the results obtained with Chleusebairgue's and other motor driver tests, with auditory tests for asdic operators, and with the physical agility test in the Army, have already been quoted. The latter (S.P.16) was measured by the time taken to transfer a set of steel rings from two upright posts to two others standard distances apart, the mean time being approximately one minute. The test was devised to resemble the duties of gun numbers and infantry, and did yield significant but small correlations with proficiency among the latter (averaging ·22). But it was very unreliable being much affected by the floor surface and the recruits' shoes, by their state of health at the time, and by the amount of enthusiasm the tester was able to stimulate. Even this test was found to depend to some extent on *g*. It had a certain propaganda value, but very little else. Several dexterity tests were tried out in mechanical branches of the Navy, Army and A.T.S., with no success. A finger dexterity test (Co-ord.-C) which is being used in the R.A.F. has been described by Cockett (1947).

Considerable research was devoted to tests for gunnery though it was hampered by the extreme unreliability of the available criteria. Instructors and officers could not say which recruits really were good or poor gunners, and objective proficiency tests were ruled out by lack of time and ammunition. Two miniature situation tests—Hotoph's layers test and Craik's predictor—did give promising results, but it was impracticable to construct and maintain the apparatus needed for testing, say, a 1,000 men a week. Among A.T.S. recruits in anti-aircraft batteries, a set of nine sensory-motor tests gave correlations ranging from +·165 to −·106 with assessments of proficiency at a practice camp, whereas the Matrices test correlated ·323 with the same criterion, and Table XXX shows that group paper-and-pencil tests yielded even better predictions in later follow-up. In the Navy it was observed that marks awarded by the gunnery officer during the short gunnery course at seaman training establishments were as predictive of later success at a gunnery school as was T2 (correlations of ·47). Of forty-three men receiving recommendations at their seaman establishment only one (2·3 per cent.) failed, whereas of 162 not recommended, 19·7 per cent. failed. This suggests that paper-and-pencil + work-sample testing constitutes a better approach to the selection of gunners than aptitude testing.

Chapter XVI shows that the same conclusion is probably true of aircraft pilots. However, two complex co-ordination tests, devised by the Cambridge University Psychology Department, were found to possess some validity. S.M.A.3 (Co-ord.-A) requires the testee, seated in a mock-up cockpit, to keep a moving spot of light as near as possible to the centre of a screen, during three runs of 1½ minutes each. He can adjust its vertical position by a control column, and its horizontal position by a rudder bar, so that gross co-ordination of both hands and feet are tested. The runs are scored by the amount of time the light falls outside a central square. The control of velocity test (Co-ord.-B) consists of a rotating ivorine cylinder, marked out as a winding road with "kerbs" and "obstacles" punched in it. The testee steers a bronze ball along this road by a hand steering wheel. Each time he fails to avoid an obstacle an electrical counter is operated. The wheel is so geared to the ball that motor anticipation is essential, since rapid movements at the last moment cause it to wobble violently. The U.S. Army Air Force, with its much greater resources of personnel and materials, was able to develop several other apparatus tests possessing useful validity for pilot selection. Nevertheless, its paper-and-pencil tests generally gave better predictions (cf. Flanagan, 1946; Davis, 1947). Similar work was done by Biesheuvel for the South African Air Force.

In yet another experiment on naval radar operators, two tests were found to add appreciably to T2 in the prediction of training marks. The Echo test, devised by Craik, provided a mechanical representation of the "scan" of a commonly used radar set and was scored by means of the smallness of the echo, or peak in the tracing, which the testee could place correctly. This had to be abandoned, however, since no method could be found for converting it into a group test. The other test, Pointers, required testees to gauge the positions on a scale to which two arrows were pointing. Forty such items were shown on large cards at an increasing rate, the object being to observe the testees' reactions to stress, and their tendency to breakdown or "flap." Whether this really acted as a test of temperament could not be determined, but it certainly depended to a considerable extent on mathematical ability. Hence the scale and graph reading test, mentioned above, was developed instead.

Morse Aptitude and Auditory Tests

The Morse Aptitude test (S.P.10) used for selecting Army and A.T.S. signallers during most of the war consists of seventy-eight pairs of sound patterns, presented by gramophone and headphones. Testees judge whether the two patterns in each pair are the same or different. This is an American test—the U.S. Signal Corps Code Aptitude test—dating from 1918. Since its reliability is inadequate, American psychologists sometimes trebled its length, but later discarded it in favour of morse learning tests. The U.S. Army "Speed of Response" test has now been adopted by the Royal Navy, and was proved in an Army investigation to possess superior validity. It, too, is given by gramophone records. The testees first get instruction and practice on three actual morse patterns for seventeen minutes, and are then tested for the quickness and accuracy with which they can receive these patterns, for about ten minutes. Naturally recruits with previous experience have an advantage, hence it is supplemented with a short record for testing receiving ability at four and six words a minute.

The old Morse Aptitude test occasionally gave correlations of ·30 to ·35 with receiving proficiency after training, but in extended follow-up its validity was lower and could be attributed largely to its dependence on intelligence. Correlations of Arithmetic, Clerical and Spelling tests with proficiency were usually superior. It was particularly defective in that its correlation was better at the top than at the bottom of the scale, i.e. it was fairly efficient in differentiating potentially very good from average telegraphists, but almost useless at picking out recruits who could never learn morse. Even the new morse learning test, though claimed to have a validity coefficient of ·49 in the American Forces, depends to a still greater extent on intelligence and does not appear to give good predictions beyond the early stages of training. Moreover, it was found in the Navy and R.A.F. that speed and accuracy of receiving fluctuate erratically during training, and even from day to day. For example, accuracy at the sixth, tenth, fifteenth and twentieth weeks gave the following correlations with final accuracy at the twenty-fifth week (20 words/min. rate)—·51, ·67, ·65 and ·71. Several trainees with very poor accuracy of under 60 per cent. at seven weeks (8 words/min.) eventually passed with 90 per cent. or over*. Still greater alterations occurred among telegraphist air gunners whose training lasted forty-two weeks, and among Army boy tradesmen. Thus it hardly seems likely that any test lasting only half an hour can be expected to give good predictions of eventual efficiency.

A group auditory acuity test (S.P.19) was applied for a period in the Army and A.T.S. in the selection of signallers and certain categories of operators. Here a series of numbers is dictated by gramophone and headphones, with decreasing intensity. The reliability of the test was hopelessly inadequate, and in only one group—A.T.S. special operators—was it ever shown to have any validity (a corrected correlation of only ·24). Since this test has been adopted by several Education Authorities for identifying hard-of-hearing children, it is to be hoped that careful investigation of its value will be made in schools before any reliance is placed on its results. Two gramophone tests developed in the R.A.F. are briefly described by Dickson ("Brit. Med. Bull.", 1947). One involves counting the numbers of pure-tone pips sounded at decreasing levels of intensity; the other requires recognition of words heard through a background of engine noise.

---

* The reliability is doubtless lowered by the inadequacy of the methods of testing proficiency. However, in R.A.F. investigations the American objective Code Receiving tests were employed, and highly significant individual variability was still found.

Selection of Instructors

Investigations were made in all the Services into the selection and training of instructors, those in the R.A.F. being the most extensive. Youthfulness, intelligence and education were found to be of considerable importance. Thus the Mathematics test 3b correlated ·45 with assessments at the end of training among 161 naval gunnery instructors; and among driving instructresses in the A.T.S., instructions, spelling, the "Mec" test and previous education yielded good results. But a particularly effective miniature situation test was developed in the latter research. This consisted in giving the candidate a Meccano model to study, and then having her explain to a "stooge" how to make the model, given only the separate Meccano parts. Ratings by a psychologist who listened to the explanations of 146 candidates were considerably more predictive than ratings, based on interview only, by a P.S.O., a psychiatrist, and a mechanical transport officer. Further work confirmed the value of this method and demonstrated that it could be applied equally successfully by specially trained P.S.O.s. But other, more objective, tests which were tried out as alternatives to this subjective rating showed no promise.

Personality and Temperament Tests

In Chapter IV we described the importance attached to personality qualities among naval and Army officer candidates, together with the interview and group task methods developed for assessing them. It would have been extremely useful to possess reliable tests which could be applied to all recruits, both for screening cases with suspected neurotic or psychopathic tendencies (who could be sent for interview by psychiatrists), and for judging leadership, industriousness, and the like, which were clearly of the greatest significance in all naval and military employments. Even if the worth of the W.O.S.B. techniques had been more fully established, it would have been impracticable to extend them to very large numbers. Hence the assessment of personality was based almost entirely on P.S.O.s' interviews, whose inadequacy we have already admitted. The difficulties of personality testing have often been described, and these are enhanced when numbers are so large that the tests must be group ones, or—if individual—very short, and when a considerable proportion of the testees are barely literate, and of the testers only moderately skilled. Little time could be devoted to the problem by qualified psychologists; thus, we cannot claim to have advanced very far.

Actually a greater amount of relevant work was carried out by Eysenck (1947a) and his collaborators at the Mill Hill Emergency Mental Hospital, and some of his main methods and results will be summarised first. By means of factorial analysis, experiments and tests, chiefly with neurotic Army and A.T.S. patients, he was able to confirm the existence of two major dimensions or factors of personality, comparable to *g* and *v : ed* v. *k : m* in the field of abilities. These are:

(1) General neuroticism, or stable and integrated *v.* maladjusted and poorly-organised personality.

(2) Extravert *v.* introvert tendency, the extremes of which are represented by hysteria and dysthymia (anxiety or obsessional neurosis) respectively.

A number of tests were found to differentiate to some extent between hysterics and dysthymics, or between maladjusted and normals. The latter included:

(i) A personality questionnaire containing such questions as:

Have you ever been off work through sickness a good deal? Yes. No.

Did you find it difficult to make friends? Yes. No.

This was presented as a "medical" questionnaire, in order to make the hypochondriacal questions more acceptable. But it also contains items bearing on inferiority feelings and lack of sociability, i.e. characteristics sometimes regarded as introverted. Eysenck finds that these characteristics are associated with neuroticism (in his sense of the term) rather than with dysthymia, and the questionnaire studies surveyed by Vernon (1938a) support him. The mean scores on forty questions among 300 normals and over 500 neurotics were 2·6 and 19·5 respectively.

(ii) Body Sway. (*a*) Static Ataxia. The testee stands upright with his eyes closed, and a thread is attached to the back of his collar, and connects with a pointer which records the amount of sway on a smoked drum. Among 120 normals, none swayed more than 2 inches in 30 seconds, whereas 34½ per cent. of 900 neurotic swayed 2 to 6 inches or more.

(*b*) Suggestibility. A gramophone record is now started which reiterates the suggestion: "You are falling, falling forward. . . ." Sway is generally increased and some 60–70 per cent. of neurotic patients now move more than 2 inches, but only 5–10 per cent. of normals do so.

(iii) Dark Vision. The Livingston rotating hexagon test, used in the R.A.F., showed that neurotics tend to have much poorer dark adaptation than normal recruits.

(iv) Rorschach Inkblots (Harrower's multiple-choice group test). The ten inkblots are presented by slides, each with nine possible responses, some commonly chosen by neurotics, some by normals. The testees rank these in order of appropriateness, and the sum of their rankings of the neurotic responses constitute their scores.

(v) Word Association (adapted from Malamud, 1946). Two possible associations are supplied for each of fifty words, one characteristic of normals, one of neurotics. The score is the number of neurotic responses preferred.

(vi) Perseveration Tests. It was confirmed that normals tend to obtain moderate scores and neurotics very high *or* very low scores on ordinary motor perseveration tests.

(vii) Persistence. In one simple test (similar to that of Fernald, 1912), the testee sits on a chair and keeps the heel of one shoe about an inch above the seat of a second chair as long as possible. The average times for hysterics and dysthymics were fourteen and thirty-one seconds, and for normals over one minute.

(viii) Personal tempo or normal speed of work on such tasks as manual dexterity or co-ordination tests was found to be slower among neurotics. Good results were obtained with the O'Connor Tweezer Dexterity test and the Track Tracer, where the subject traces a path with a metal stylus between rows of holes on an ivorine sheet, and a buzzer sounds each time a hole is touched.

Eysenck has advanced some way towards the production of reliable batteries of tests for measuring neuroticism and hysteria-dysthymia, but not all the tests mentioned would be suitable for large-scale application, and evidence regarding the validity of some of them is as yet hardly conclusive.

Indirect Tests

One of the difficulties of personality testing is that, when subjects realise the object of the test (or if they misinterpret the object) they will always tend to respond so as to put themselves in the most favourable light. Hence for many years both Burt and the present writer have advocated a more indirect, qualitative approach, where the subject reveals his personality or temperament by the manner in which he tackles various tasks—usually tests of abilities. This was the principle adopted by Biesheuvel in his work with the South African Air Force, not yet published. He was able to show that carefully trained testers could make consistent judgments on specially prepared rating scales, of qualities revealed by the subjects' methods of approach, reactions to difficulties, etc., and that such judgments had useful predictive value in the selection of air crew. Moreover, several aspects of performance at sensory-motor or other practical tasks, which could be objectively measured, were found to be temperamentally significant. When, however, similar methods were tried out in the selection of U.S.A.A.F. pilots, the results were generally disappointing (cf. Davis, 1947; Guilford. 1948).

Another approach to the investigation of temperament was adopted by the Cambridge Unit of Applied Psychology. It was considered that temperamental stability or tendency to "flap" may be shown by the *trend of performance* in tests which impose prolonged stress on the subject. One of these was the Pointers test (p. 249), another the Track Tracer, applied at maximum speed for several minutes. A third was a modification of the McDougall-Schuster dotting machine, where the holes which the subject tries to hit with a stylus revolve past him at a constant rapid rate. The performances of the subjects on these tests in successive half-minute or other periods are recorded, and from the upward or downward trend, or the irregularity, a somewhat subjective rating of temperament is reached. Adequate evidence as to the validity of these methods is not yet published, and they would hardly appear suitable for large-scale use, both because they involve a fairly lengthy session of individual testing, and because they could scarcely be entrusted to semi-trained testers. For the same reasons no attempt was made in the Forces to apply projection tests except to relatively small groups such as officer candidates, although the original Rorschach inkblots, thematic apperception, and other projection tests have come to be regarded, in recent years, as the most promising techniques for the diagnosis of personality. It was noted too, that tests of the projection type tend to be unpopular both among testees and employers (e.g. Army officers) because of their apparent lack of relevance to the job. The psychologist cannot afford to disregard this factor of "face validity" completely.

Another test providing measurable indications of such traits as calmness, foresight, initiative and persistence is Cattell's (1941, 1944) C.M.S. or cursive miniature situation test, which is an elaboration of the dotting test. He gives striking evidence of differentiation between delinquents or other unstable personalities and normals. Though it failed to live up to its promise in a small trial with W.O.S.B. candidates, this may have been due to our inability to construct suitable apparatus and dotting sheets. It suffers, however, from a serious drawback, namely, that the scoring is extremely tedious and time-consuming. Only if this could be simplified or automatic electric scoring introduced, would it become a practicable test. As mentioned above, it is possible that the Clerical or Instructions test is another indirect measure of certain temperamental qualities, comparable to Biesheuvel's, the Cambridge, and Cattell's tests.

Tests Tried Out in the Army

As already indicated, no precise data are available on the value of the projection tests used at W.O.S.B.s. An early experiment, however, revealed significant differences between the grammatical forms of response to the group word association test among fifty company commanders, fifty W.O.S.B. passes and fifty fails (of equivalent intelligence level). For example, the mean percentages of complete sentence responses were 41, 31 and 19 per cent. respectively. No simple method of scoring either the form or content of responses could be devised, which could be used by unskilled testers, except the number of blanks or failures to respond. This was tried out at an Army selection centre on 218 recruits, who were interviewed by psychiatrists and assessed for temperamental stability *v.* neurotic tendencies on the Culpin scale. No correlation whatever was found, and failures to respond appeared to depend chiefly on illiteracy and on the testing conditions (e.g. the particular tester). By contrast a "medical" questionnaire, similar to Eysenck's but based on Part III of the Cornell Selectee Index (Weider, 1944), gave an agreement represented by a tetrachoric correlation of ·31 with the Culpin scale ratings, which is moderately promising.

The group Rorschach (ranking) test was given to 200 A.T.S. in a selection centre, fifty of whom had been picked out as having personality problems by the P.S.O. and psychiatrists, 100 as intermediate, and fifty as being the most stable and well adjusted in the group. The median scores were 215, 223 and 234 respectively, corresponding to a correlation ratio of ·26 (note that a *low* score indicates instability). If a borderline mark is fixed at 200, it would cut off 25 per cent. of the most neurotic and 12 per cent. of the remainder. Such differentiation is far from adequate, but it should be pointed out that the testees were a selected group. In general recruits are not sent to a selection centre unless they are in some degree maladjusted, yet, at the same time, few are likely to be in a serious neurotic state. Thus this result does not conflict with that of Eysenck who obtained median scores of 205 and 231 among neurotics and normals. The test was found to depend to some extent on *g* and *v:ed* (giving correlations of ·33 with Instructions and Spelling). However, with the revision of the responses presented to the testees, it might in combination with other tests be useful for screening purposes. A new inkblots test, suitable for group application, was prepared by Harvey, but has not yet been validated.

It may seem curious that neurotics should differ from normals on sensory and motor tests such as dark vision and dexterity, though not on tests of mental abilities. But Slater (1944) claims that temperamental inferiority would be expected to be associated with constitutional physical inferiority, and has shown that significant differences occur on other tests such as Agility (S.P.16) and visual acuity. For example, a simple acuity test was given to 169 neurotic patients and to 2,233 normals (none of whom wore glasses), and only 30 per cent. of the neurotic group reached or exceeded the normal median score.

Some experiments were carried out on gradings of recruits by their own fellows, which appeared to show that they were capable of nominating promising candidates for commissions, who had not been picked up by the usual channels. Unfortunately, this technique was received with some suspicion. Subsequent American work has shown that gradings by fellow cadets at officer training schools give quite good predictions of success in battle a year or two later (Baier, 1947). The correlations of ·42 and ·51 are distinctly more promising than correlations obtained for W.O.S.B. or for O.C.T.U. grades. This suggests that it might be worth taking fellow employees' opinions into consideration in managerial and other civilian selection.

Personality Inventories and "Medical" Questionnaires

Evidence regarding the validity of these questionnaires in civilian practice is decidedly unfavourable (cf. Vernon, 1938a; Ellis, 1946), and one would naturally expect them to be more unsuitable with low-grade, semi-illiterate recruits. Moreover, they tend to arouse suspicion among naval and military officers. Nevertheless, they achieved surprisingly good results both in America and in this country, especially when given with medical backing. Conrad (1947) suggests that this was due partly to the greater heterogeneity of recruits than of most civilian samples, and partly because the questions asked are very similar to those which psychiatrists ask when they are assessing the neurotic tendencies of the testees. Possibly, too, average and sub-average men and women accept the questions more readily than sophisticated university students, on whom much of the earlier work was performed. Yet another factor is that civilians usually try to make themselves out as normal as possible, whereas recruits may adopt the opposite attitude.

Two main types of validatory criteria have been employed, neither of them very satisfactory:

(i) Assessments of neurotic tendencies in a miscellaneous group by psychiatrists.

(ii) Differentiation between persons who have already developed neurotic breakdown and others who are presumed to be normal. Naturally, it does not follow that the neurotics would have given the same answers before they reached the stage of hospitalisation or discharge.

Nevertheless, one American study demonstrated the value of the National Defence Research Council's personality inventory in differentiating recruits who, in the following year, were discharged or committed offences, or were promoted. A long-term psychiatric follow-up of the Bennett-Slater questionnaire and the Harvey designs (inkblot) test has been undertaken in the British Army.

While adaptations of the Cornell index and the N.D.R.C. inventory (short format) were "got across" to recruits in this country quite effectively, the most successful of such tests was that published by Bennett and Slater (1945), where the neurotic answers are ingeniously concealed. This is in ten sections, most of which gave significant differences between neurotic and surgical patients at Sutton E.M.S. Hospital. The first three consist of ordinary questions, in about half of which a positive answer indicates neurotic tendencies, in the remainder a negative answer. Thus the testee who wishes to misrepresent himself cannot merely answer "no" throughout. The sections deal with symptoms of anxiety, hysteria and depression, respectively. The middle one was the least satisfactory. The next four consist of lists of "annoyances," testees checking each item they find annoying. These are classified as follows:

(1) Frustration of self-assertion, e.g. "Somebody tells you how to do your job."

(2) Personal inadequacy, e.g. "You forget what you're looking for."

(3) Dirt or untidiness, e.g. "An unmade bed."

(4) Noise, e.g. "The sound of hammering."

Nos. (1) and (3) are regarded as things which might annoy any normal person, but Nos. (2) and (4) are much more likely to affect neurotics. All types are mixed in the test blank, and the score is based on the difference between them.

The last three sections are revisions of Pressey's cross-out test, where the subject crosses out items:

(1) For which people should be blamed, e.g. "Flirting, speeding."

(2) Which they have worried about, e.g. "Loneliness, falling."

(3) In which they are interested, e.g. "Football, comedians."

In the first two sets items are chosen as likely to affect neurotics, and in the third as likely to appeal more to non-neurotics.

Slater shows that when the sections are appropriately weighted a certain borderline score will cut off 61 per cent. of neurotics and only 2 per cent. of normals. This needs, of course, to be confirmed on another, larger, group. A defect of the test is that it takes nearly an hour to give. Moreover, the scoring is so lengthy that it could never be used as a routine test. However, some of the sections and some of the items are more diagnostic than others, and a more practicable revision might well be constructed. American experience suggests that a short twenty to thirty item test is as effective as a more elaborate instrument*.

Another questionnaire tried out successfully on motor drivers (cf. p. 230), contained fifty miscellaneous questions such as:

1. Can you ride a bicycle?
2. Do you think all cars should be restricted to 30 m.p.h.?
3. Do you always take a torch with you in the blackout?
4. Do noise or vibration in a bus or train ever give you a headache?
5. Do you prefer indoor to outdoor games?

All of these differentiated good from bad drivers, perhaps for a variety of reasons. No. 1 may indicate "machine sense." "Yes" answers to Nos. 2 and 3 suggest undesirable nervousness, while Nos. 4 and 5 are of the conventional psychosomatic or neurotic type. Such questions are less embarrassing than those in most personality inventories, and their object is not at all obvious—some of the "good" answers being "Yes," some "No." Though the scoring is relatively easy it was the time taken over this and over giving the test which prevented it from being more widely used.

Interest Blanks

Strong's Vocational Interest Blank and Kuder's Preference Record have been shown, in America, to give useful predictions of suitability for a number of professional and commercial careers. Probably they are much less effective among average and low-grade adults, for at these levels the responses to interest items depend to so large an extent on intelligence and education, or else on temporary fashions. For instance, the interests ticked by a skilled instrument mechanic might resemble more closely those of an equally intelligent clerk than they would those of a lower-grade metal worker, say a welder. Nevertheless, a test consisting of a list of jobs and leisure-time interests was standardised in the Army on recruits representative of the various training recommendations—drivers, mechanics, clerks, etc. Thus any new recruit's interests could be scored for their resemblance to those of men in each of these types of employment. Though the scores had poor reliability, they showed useful predictive value in several experiments, and the test was abandoned mainly because of the time required for scoring. A test of this kind might have been useful with a higher-grade population such as officer candidates, but preliminary studies with passed and failed candidates in the Navy and Army did not reveal many consistent interest differences. The same finding was reached in work with U.S.A.A.F. pilots, and it was concluded that aviation and other interests are better assessed by information tests (Guilford, 1948). Probably, as Wilson (1945) suggests, the type of interest is of less importance than the seriousness with which it is pursued, and it is difficult for a pencil and paper test to elicit this attribute for a large number of interests.

---

* Cf. Stuit (1947). This book contains an excellent discussion of the construction and uses of personality inventories.

Conclusion

Our conclusions regarding the value of specialised aptitude tests in peace-time have already been indicated in Chapters X and XII. How about personality tests? Clearly the prospects for large-scale application are not very bright. When ample time and skilled testers are available, and the number of testees is small, objective testing of general neuroticism and other "dimensions" is possible by batteries such as Eysenck's. Trained psychologists, too, can make valuable personality diagnoses by projection tests and by observation of "manner" of performance, etc. For routine use by less qualified personnel, it seems that questionnaires could be constructed along the lines of Bennett-Slater's and the motor drivers', though they would need to be re-validated under the very different conditions of motivation present among applicants for civilian employments. Multiple-choice projection tests also certainly deserve further exploration. Apart from these the most hopeful line for assessment of personality would appear to be the development of better interviewing techniques, which could be partially standardised among different interviewers.

CHAPTER XVI

MAIN R.A.F. SELECTION FINDINGS

Abstract.—This Chapter gives detailed information on the following topics:

1. Pilot training wastage before the introduction of grading.
2. Relation between speed to solo and training achievement.
3. Quarterly pilot training wastage from 1940 to 1943.
4. Relation between grading score levels and elementary flying pass rates.
5. Layout of the original grading score card.
6. Use of the raw grading scale by a number of different schools.
7. Relation of grading performance to subsequent accident rate in pilot training.
8. Relation of grading to age differences and failure rate.
9. Composition of the original (April, 1944) air crew aptitude test battery.
10. Reliability of tests in the aptitude test battery.
11. Validation of aptitude test category batteries against different training stages.
12. Interpretation of aptitude test results in relation to expressed preferences and Service requirements.
13. Personality and character traits deemed important in air crews.
14. Reliability of the above trait assessments.
15. Average intelligence test scores for selected R.A.F. ground trades.
16. Multiple correlations between G.V.K. scores and training results for a variety of R.A.F. and W.A.A.F. trades.

The purpose of this Chapter is to present the findings on R.A.F. selection methods that are most likely to be of general interest. So far as possible the Tables will be left to speak for themselves, but a certain amount of verbal clarification will usually be necessary.

Air Crew Selection

The chart (p. 263) shows what had happened after two years to air crew volunteers who had been accepted for pilot training by Aviation

Candidates Selection Boards. Out of every 100 cadets entering training there were 41 who had qualified for operations, 17 who had qualified but were engaged on non-operational duties (primarily as instructors), 6 who were still under training, and 36 who had failed to qualify as pilots. It is clear that wastage was heavily concentrated at the elementary flying (E.F.T.S.) stage which was being used as the main means of pilot selection.

TABLE XL.—RELATION BETWEEN SPEED TO SOLO AND TRAINING ACHIEVEMENT
Total cases: 9 E.F.T.S.s, 36 courses, N = 2,292
Did not solo = 554 (24%). Did solo = 1,738 (76%)

Group	*Hours Dual to Solo*	*Total Cases*	*Sample selected for follow-up*	*Flying Ability*					
				E.F.T.S.			*S.F.T.S.*		
				Supr.	*Med.*	*Inf.*	*Supr.*	*Med.*	*Inf.*
	Hrs.			%	%	%	%	%	%
V. fast	5-8	193	150	52	45	3	44	48	8
Fast .	9-10	610	150	34	58	8	41	48	11
Medium	11-14	785	150	20	66	14	29	50	21
Slow .	15-21	150	150	6	59	35	14	54	32

The selection method known as Grading presumes a close correspondence between learning speed and subsequent achievement and before it was introduced it was necessary to prove the reality of this association. Table XL which is based on the E.F.T.S. records of some 2,300 cadets during the summer of 1941 shows two things:

(i) That almost a quarter of the cadets have to be suspended without going solo at all.

(ii) That if those who do go solo are divided into four groups based on the number of hours required to reach this stage, consistent superiority is found throughout training by the very fast over the fast, the fast over the medium and the medium over the slow*.

The first of the above findings is in line with the conclusions of nearly all other Air Forces, that of every four volunteers who are medically fit and who appear in every way suitable for pilot training, one will, in fact, be discarded in war-time without even going solo.

---

* The table yields tetrachoric correlations of ·39 and ·25 between time to solo and E.F.T.S. and S.F.T.S. results respectively. The first figure would be higher if all pilot applicants went to E.F.T.S. and the second much higher if they all passed on to S.F.T.S.

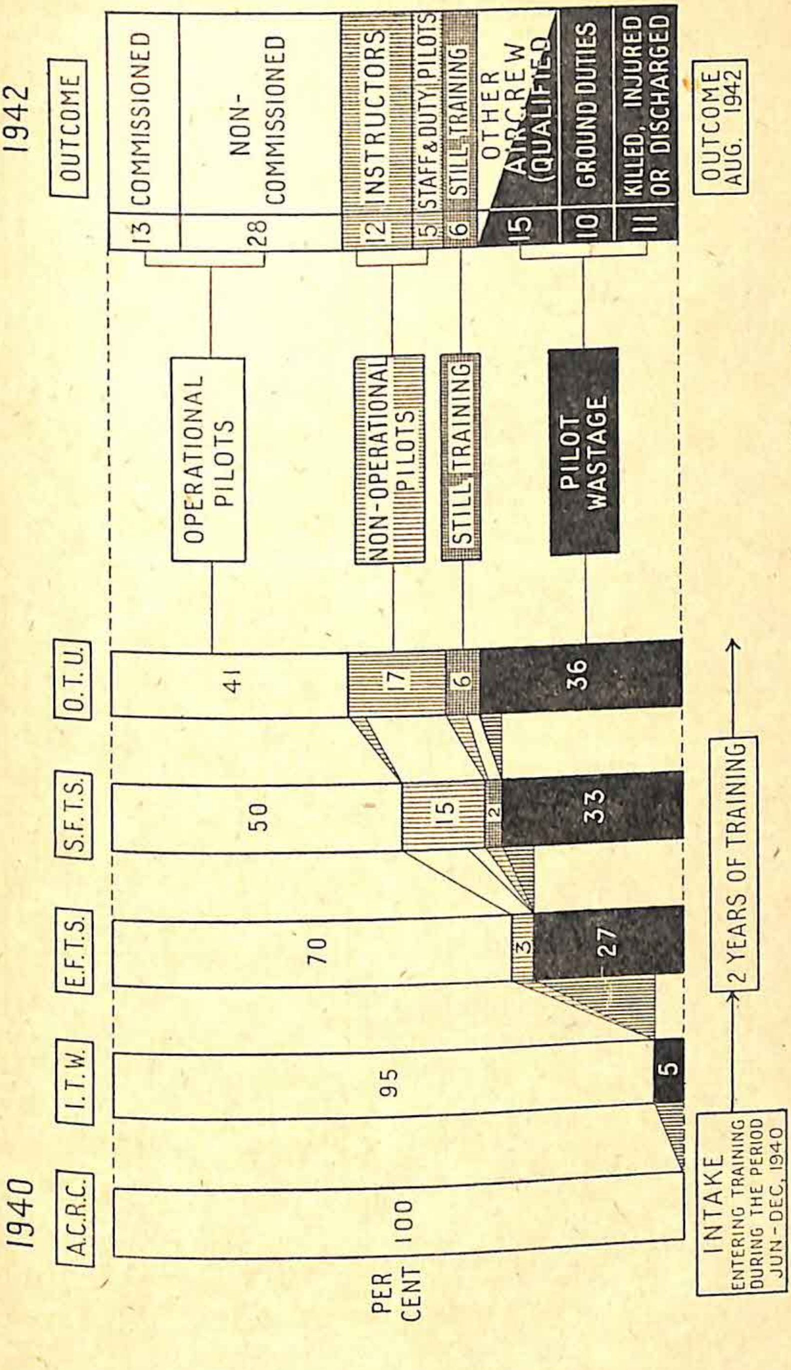


Fig. 10. Percentage of Pilot Training Wastage Before Grading.

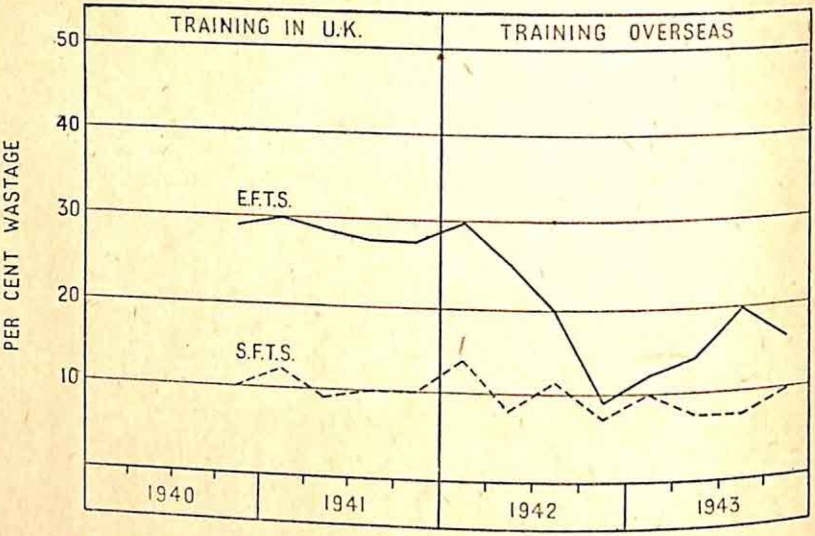


FIG. 11. PILOT TRAINING WASTAGE BY CALENDAR QUARTERS.

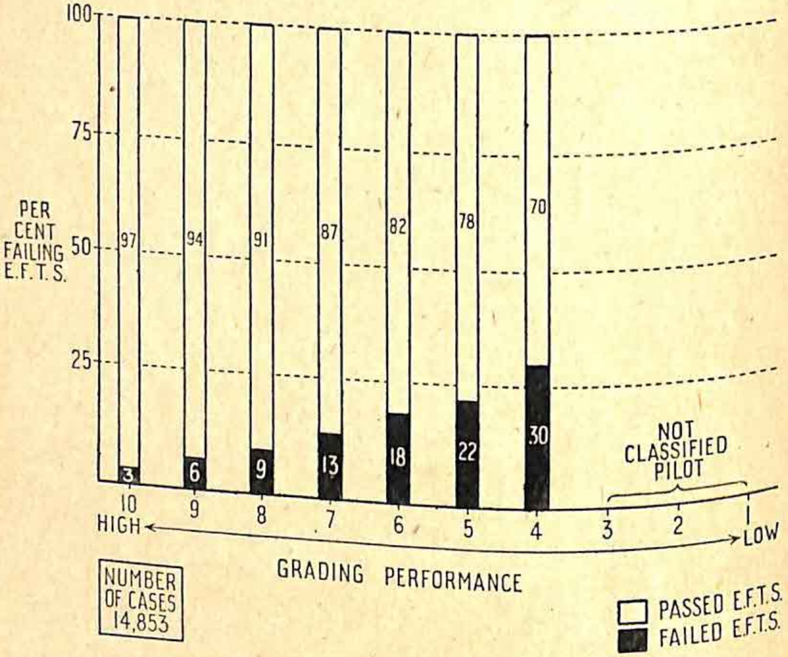


FIG. 12. FLIGHT TEST (R.A.F.) E.F.T.S.

Up to and including the first quarter of 1942 flying training schools were dealing with cadets who had been nominated as pilot trainees by selection boards using the "impression" method. This gave a gross pre-O.T.U. failure rate of about 40 per cent., when R.A.F. schools were in the U.K., rising considerably higher when they went overseas. In the second quarter of 1942 many of the cadets had received a limited amount of casual flying instruction before going overseas; the failure rate then became about 35 per cent. In the third quarter all cadets had been through the grading procedure, but draft commitments made it necessary to send nearly all overseas whether they had shown much or little aptitude for flying—under these conditions the subsequent failure rate came down to 30 per cent. By the fourth quarter of 1942, grading came into full operation, only those who had demonstrated a relatively high degree of flying aptitude going forward. This method of

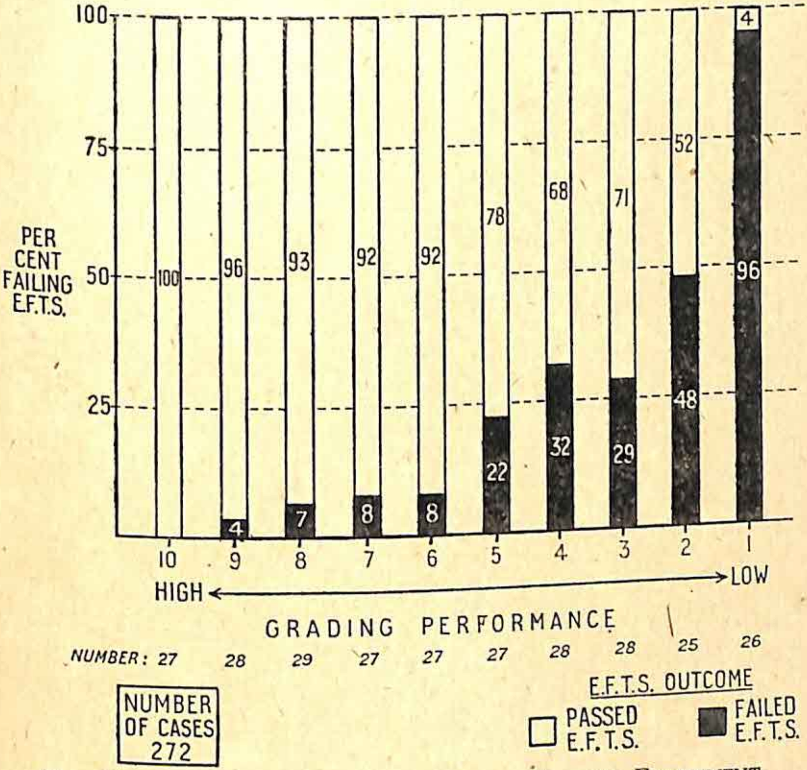


FIG. 13. FLIGHT TEST (R.C.A.F.) SPECIAL GRADING EXPERIMENT.

"quality selection" reduced the overseas failure rate abruptly to 15 per cent. The subsequent increased wastage rates in 1943 were the result of an attempt to secure higher quality by failing borderline cases who would hitherto have been passed. This policy was not successful and counter-measures had later to be taken to restore the original standards.

Figs. 12 and 13 show the relationship between degrees of grading proficiency and E.F.T.S. performance. Ordinarily no one in the lowest three grading groups would be sent forward and comparatively few from No. 4. The results in Fig. 12 are based on nearly 15,000 pilots and show a steady diminution in the E.F.T.S. failure rate as the higher grading groups are reached. Fig. 13 records similar data for the only complete grading population sent forward; the numbers are small (272 in all), but the increase in failure of Groups 1 and 2 is large enough to command attention.

R.A.F. Form 325(1)

No. . . . E.F.T.S.	GRADING ASSESSMENT FORM	Test No. 1

No. . . . . Rank. . . . . Name. . . . .

Instructors:	Dates grading { From . . . . To . . . .
Tester:	For Office Use Total Marks . . . . Extended by . . . . Checked by. . . . . Time to 1st Solo . . . . Time to Test. . . . .
Solo Before Test Solo Immediately After Test Not Solo	

1 a	*Taxying and Tarmac Check* Tarmac Check . .	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Use of Controls and Eng..	0 1 2 3 4 – 6 7 8 9 10	(1)	
c	Safety-speed and look-out .	0 1 2 3 4 – 6 7 8 9 10	(1)	
2 c	*Take Off and Climb* Pre take-off airmanship .	0 1 2 3 4 – 6 7 8 9 10	(2)	
b	Use of elevator . .	0 1 2 3 4 – 6 7 8 9 10	(4)	
c	Keeping straight . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
d	Use of engine. . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
e	A.S. & R.P.M. on climb .	0 1 2 3 4 – 6 7 8 9 10	(3)	
3 a	*First Approach* Position to commence glide	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Selection of turning point	0 1 2 3 4 – 6 7 8 9 10	(1)	
c	Maintaining correct speed.	0 1 2 3 4 – 6 7 8 9 10	(2)	
d	Adjustment of glide path .	0 1 2 3 4 – 6 7 8 9 10	(1)	
4 a	*First Landing* Selection of landing path .	0 1 2 3 4 – 6 7 8 9 10	(2)	
b	Judgment of check height	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Handling during hold off	0 1 2 3 4 – 6 7 8 9 10	(3)	
d	Quality of landing attempt	0 1 2 3 4 – 6 7 8 9 10	(6)	
e	Subsequent actions . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
		Carried Forward		

		Brought Forward		
5 *a*	*Medium Turn (Right)* Going in and looking round	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Accuracy of turn to bank .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Maintaining constant bank .	0 1 2 3 4 – 6 7 8 9 10	(1)	
d	Coming out w'out skid or slip	0 1 2 3 4 – 6 7 8 9 10	(1)	
6 *a*	*Medium Turn (Left)* Going in and looking round	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Accuracy of turn to bank .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Maintaining constant bank .	0 1 2 3 4 – 6 7 8 9 10	(1)	
d	Coming out . . .	0 1 2 3 4 – 6 7 8 9 10	(1)	
7 *a*	*Steep Turn With Engine* Going in and looking round	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Use of throttle . . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Accuracy of turn to bank .	0 1 2 3 4 – 6 7 8 9 10	(3)	
d	Keeping height . . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
e	Coming out . . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
8 *a*	*Medium Gliding Turn (R)* Increase speed before turn .	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Accuracy of turn to bank .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Maintaining correct speed .	0 1 2 3 4 – 6 7 8 9 10	(2)	
d	Coming out at correct speed	0 1 2 3 4 – 6 7 8 9 10	(2)	
9 *a*	*Medium Gliding Turn (L)* Increase speed before turn .	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Accuracy of turn to bank .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Maintaining correct speed .	0 1 2 3 4 – 6 7 8 9 10	(2)	
d	Coming out at correct speed	0 1 2 3 4 – 6 7 8 9 10	(2)	
10 *a*	*Second Approach* Position to commence glide	0 1 2 3 4 – 6 7 8 9 10	(1)	
b	Selection of turning point .	0 1 2 3 4 – 6 7 8 9 10	(1)	
c	Maintaining correct speed .	0 1 2 3 4 – 6 7 8 9 10	(2)	
d	Adjustment of glide path .	0 1 2 3 4 – 6 7 8 9 10	(1)	
		Carried Forward		

		Brought Forward		
11 *a*	*Second Landing* Selection of landing path .	0 1 2 3 4 – 6 7 8 9 10	(2)	
b	Judgment of check height .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Handling during hold off .	0 1 2 3 4 – 6 7 8 9 10	(3)	
d	Quality of landing attempt .	0 1 2 3 4 – 6 7 8 9 10	(6)	
e	Subsequent actions . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
12 *a*	*General* Alertness	0 1 2 3 4 – 6 7 8 9 10	(3)	
b	Handling and control co-ordination . . .	0 1 2 3 4 – 6 7 8 9 10	(2)	
c	Style	0 1 2 3 4 – 6 7 8 9 10	(5)	
		Total Marks		

The constitution of the flight test is most easily conveyed by a study of the scoring card, the original version of which is shown here. From this it will be seen that the test included twelve sections the majority of which correspond to specific flight manœuvres. From three to five features are assessed for each manœuvre, every feature being marked on a 10-point scale. Differential weightings (1–6) are accorded to each detail, a hundred weighting units in all being distributed among fifty items.

It had been anticipated that, despite all efforts to standardise flight test procedure, the actual marks assigned at different schools would show differences which were really due to divergent marking habits. Fig. 14 demonstrates that this in fact proved to be the case; had the schools assigned the same marks to the same sort of performance in precisely the same way, the lines joining the quartile points would (sampling errors in the basic quality of the groups apart) have been horizontal. It was also demonstrated that when the marks assigned at any one school were studied over a considerable period of time, marking standards again showed considerable change. These two sources of instability were combated by the introduction of school conversion tables. Each table was based on the performance of the *last* hundred cases to come through a given school and involved the use of a rectilinear 20-point scale. Thus a score of 20 would correspond to a performance within the top 5 per cent. of the population whatever the raw score allotted. This

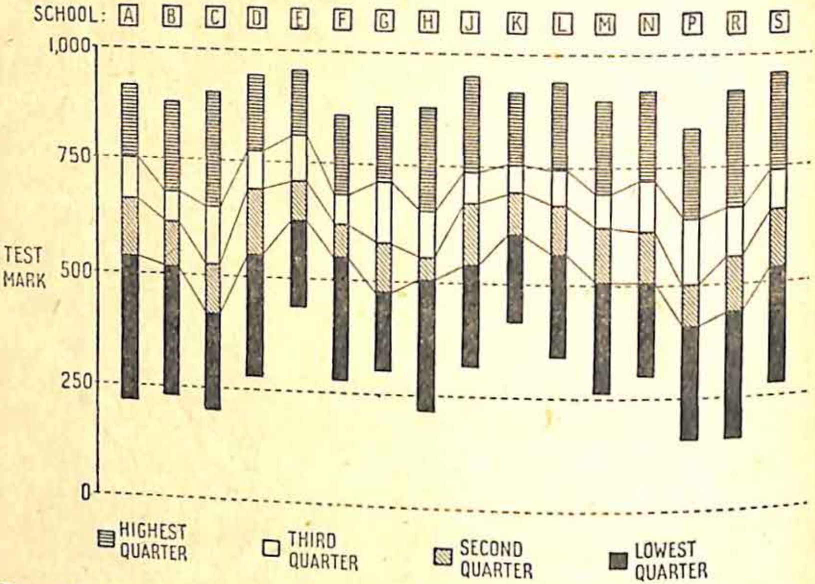


Fig. 14. Second Flight Test Marks, April, 1942. First 100 Cadets Graded at Each School.

solution was based on the assumption (demonstrated both by *a priori* reasoning and empirically) that the risk of a hundred randomly chosen cadets differing sharply in quality from that of the general population was small enough to be disregarded.

Originally equal weight was given to the two flight tests, i.e. a cadet's grading score was reached by adding the two converted figures to yield a total somewhere on a scale ranging from 40 to 2. Enquiry soon showed however that the later test was more predictive than the first. As soon as it was administratively practicable a third test was introduced with a view to strengthening reliability, and at this point the scoring was modified in two ways: decile conversion tables were brought in in place of half-deciles, and in view of its higher predictive power the last test received a double weighting.

The reliability of the individual flight tests was measured in a couple of experiments during which secondary tests, given by independent assessors, were introduced into the programme. The first of these, following immediately after the seven-hour flight test, yielded a test-retest correlation of ·765 (360 cases). The second additional test was given midway between the first and second

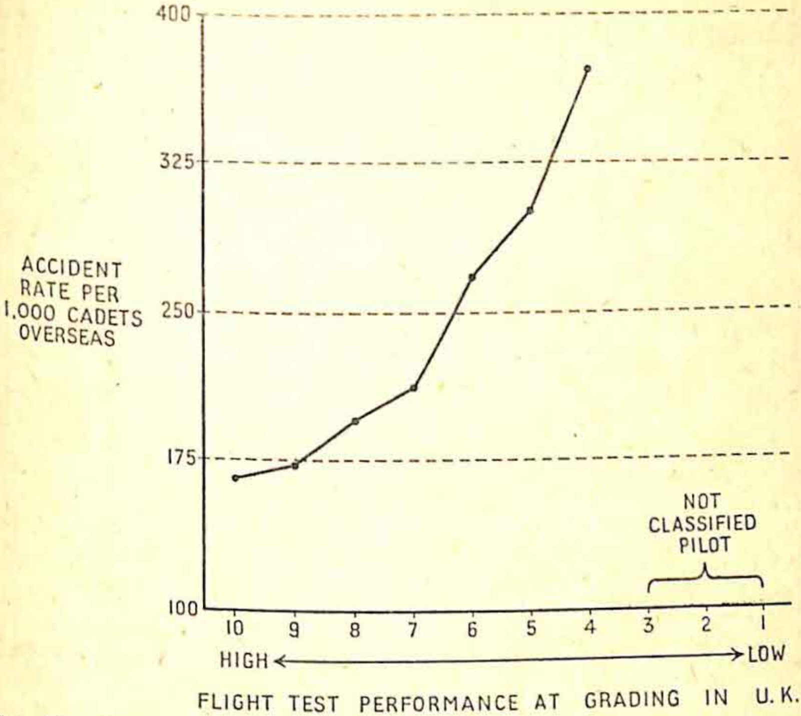


FIG. 15. RELATION OF GRADING PERFORMANCE TO SUBSEQUENT ACCIDENT RATE IN PILOT TRAINING (R.A.F. GRADED CADETS THROUGH S.F.T.S.).

flight tests so that perfect test-retest conditions cannot be claimed. The correlations (292 cases) were respectively ·676 and ·665. The relationship between the seven and eleven hour tests were also calculated on these groups, the figures being ·626 and ·561. None of these figures must be identified with the reliability of the full grading procedure which, since it then contained two tests and now has three, is almost certainly higher. Fig. 15 (based on 1,000 cases) shows that cadets who received high grading assessments subsequently showed a relatively low accident rate in pilot training, while subsequent accident rates for candidates with lower grading assessments increased sharply and steadily.

Fig. 16 shows the relationship between age and flying failure rates for both graded and ungraded populations. It might have been expected that a selection method as successful as grading has proved in other directions would have automatically levelled the failure rate for those in the different age groups who survived it. It will be seen, however, that those aged 29 and over who were

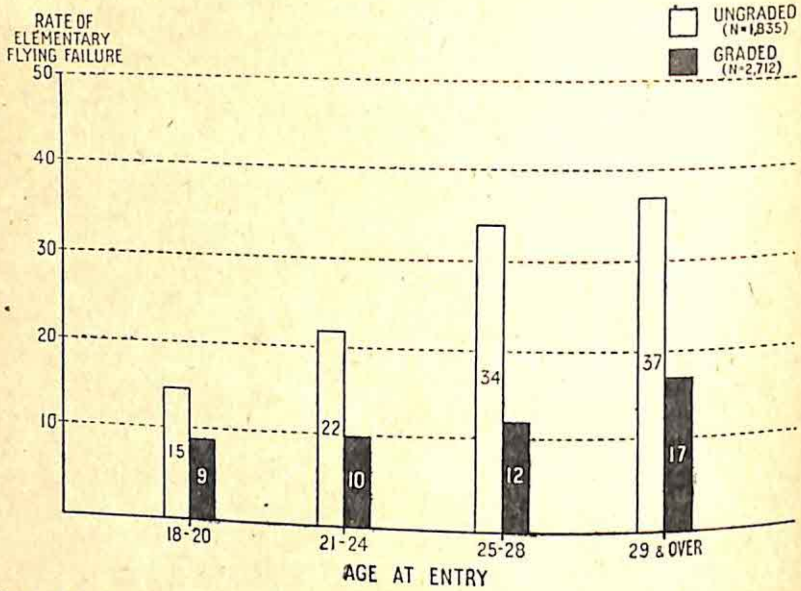


FIG. 16. RELATION OF GRADING TO AGE DIFFERENCES AND FAILURE RATE.

accepted for pilot training as a result of grading had nearly twice as high a failure rate as the lowest age group, and it therefore appears that grading does not sufficiently penalise cadets at the older ages. As, however, the R.A.F. does not normally take men of this age for pilot training, this unsolved problem is now of no practical account.

TABLE XLI.—AIR CREW APTITUDE TEST BATTERY (APRIL, 1944)

Test Description	*Short Name*	*Source*	*Pilot*	*Nav.*	*Bomber*	*W/O*	*F/E*	*A/G*
1. General Intelligence G.V.K.	Gen. A	Oxford Univ.	GENERAL					
2. Educational Attainments	Gen. B	Air Ministry						
3. Language and Judgment	Gen. C	Air Ministry						
4. Aviation Information	Gen. D	U.S.A.A.F.	EXPERIMENTAL					
5. General Maths.	Mat. A	U.S.A.A.F.		x		x	x	
6. Mathematical Reasoning	Mat. B	U.S.A.A.F.		x	x			x
7. Approximations	Mat. C	U.S.A.A.F.		x				x
8. Speed in Calculation	Mat. D	U.S.A.A.F.		x				x
9. Elementary Calculations	Mat. E	Air Ministry						
10. Table Reading	Mat. F	U.S.A.A.F.		x	x		x	
11. Mechanical Comprehension (Pictures)	Mec. A	U.S. Navy	x			x	x	x
12. Mechanical Comprehension (Diagrams)	Mec. B	Air Ministry from U.S.A.A.F.	x				x	x
13. Technical Information	Mec. C	R.A.F.				x	x	
14. Dial Reading	Ins. A	U.S.A.A.F.	x	x	x	x	x	
15. Instrument Comprehension	Ins. B	U.S.A.A.F.	x					x
16. Aerial Photographs	Obs. A	U.S.A.A.F.	x	x	x			
17. Map Reading	Obs. B	U.S.A.A.F.	x	x	x			
18. Aircraft Silhouettes	Obs. C	U.S.A.A.F.	x					x
19. Pilot Co-ordinator (S.M.A.3)	Co-ord. A	Central Medical Establishment (R.A.F.) and Cambridge Univ.	x					x
20 Control of Velocity (C.V.T.)	Co-ord. B	Cambridge Univ.	EXPERIMENTAL					
21. Finger Dexterity	Co-ord. C	U.S.A.A.F.			x	x	x	x
22. Turret Manipulation	Co-ord. D	Air Ministry	EXPERIMENTAL					
23. Morse Record	Morse A	U.S. Navy (modified)				x		

Table XLI shows the content of the two-day testing programme introduced in April, 1944 to yield aptitude measures for each of the air crew categories. The columns on the right show the tests originally used to elicit these category indices. The source column makes clear the extent of our indebtedness to the United States Army Air Force. Two of the experimental tests (Gen. D and Co-ord. B) have since been validated and introduced into the pilot battery. For details of length, timing and reliability the reader is referred to Appendix II.

TABLE XLII.—VALIDATION OF CATEGORY INDICES

Category	*Criterion*	*No.*	*Significant (·01 level) at*	*Correlation*	
				Uncorrected	*Corrected*
PILOT	Grading	923	·10	·262	·474
	Initial (Ground) Training	811	·104	·285	·473
	No Basic Training Results to hand				
NAVIGATOR	Initial Training	609	·12	·475	·636
	Basic Training	294	·18	·511	·687
AIR BOMBER	Initial Training	229	·20	·169	·229
	No tested subject came through the basic course of this now obsolete trade.				
WIRELESS OPERATOR	No meaningful training criterion.				
FLIGHT ENGINEER	Initial Training	702	·11	·487	·502
	Technical Training (Phase 1)	437	·14	·197	·280
	Technical Training (Phase 2)	396	·15	·224	·326
AIR GUNNER	Initial Training	473	·14	·365	·414
	No Basic Training Results to hand.				

To reach the different category indices the raw scores for each test were converted on to a 9-point (Stanine) scale. Twenty weighting units were differentially distributed among the tests in each sub-battery so that all the resultant indices were expressed on a common scale with theoretical range from 180 (9 × 20) to 20 (1 × 20). As a number of the tests had not formerly been tried out on R.A.F. personnel the initial weightings had to be based on evidence from the countries of origin. An all-through validation plan was drawn up as soon as the battery was brought into action and the information yielded by the very intermittent post-war training programmes is condensed in Table XLII. The corrected correlations were reached by applying univariate correction; the corrections are naturally largest where the populations selected for a category are homogeneous (as with pilots and navigators) and smallest where the selected population approximates to the basic population (as with flight engineers). It will be seen that all the corrected correlations are highly significant while all but one of the uncorrected figures (that for air bombers) is significant at the ·01 level. This single exception is of no great concern as the navigator index was always applied as a preliminary screen for the air bomber trade and the uncorrected figure yielded by this index for the same population exceeds the critical point.

It will further be noted that the figures for the navigator category are considerably greater than for any other. This is probably because intelligence and educational attainment are highly relevant for navigators whereas these general abilities (granted a certain minimum) are of relatively little account elsewhere. Where this latter situation occurs low but significant correlations for tests of a more specific character naturally acquire an importance they cannot otherwise claim.

Lack of staff has hitherto made it impossible to carry out much development work both on the tests individually and the battery as a whole. Thus the size of the battery still suggests something in an experimental stage, and there is little doubt that it could be reduced to a compacter form if more labour could be devoted to research. The same limitation has prevented many more detailed investigations particularly in the field of item analyses.

Table XLIII—Pocket Classification

Candidate	*Nav.*		*W/O*		*F/E*		*A/G*	
	Index	*Pref.*	*Index*	*Pref.*	*Index*	*Pref.*	*Index*	*Pref.*
1. Green . .	155	B	122	D	137	C	149	A
2. Brown . .	136	A	99	B	117	C	146	D
3. Grey . .	96	A	105	B	101	C	109	D
4. White . .	109	B	79	A	98	C	102	D

Table XLIII is intended to give an idea of how testing results were finally interpreted. A classification involving only four people is postulated (weekly intakes in 1944 frequently exceeded 1,000). The guiding principle has always been: "Preference granted if aptitude and Service needs permit." In this case Service needs may be taken to be for one navigator, one wireless operator, one flight engineer and one air gunner. Preferences are indicated by letters (A first choice, B second, etc.) and indices by numbers. Green, in spite of his very high navigator index, will become the gunner by virtue of his preference for this category. Brown with an A preference and a high index will become the navigator. Grey's aptitude is too low for his A preference, but just adequate for his B, so he becomes the wireless operator. The berths for White's A and B preferences are already filled and his aptitude for C (flight engineer) is just adequate (had it been much lower it would have been necessary to recommend leaving this berth unfilled). A complication arose in the classification of pilots for whom grading constitutes a secondary screen. This was met by giving an alternative classification to which approximately the lower half on the grading tests reverted.

So far comparatively little work has been undertaken on the problem of personality assessment for air crew. Throughout the war little could be done to standardise interviewing procedure, but late in 1945 a first step was taken with the issue of a questionnaire to 362 highly experienced air crew representing all commands and categories. The purpose of this was to find (*a*) what degree of agreement existed as to the relevance of different traits for air crew, (*b*) whether there was any correspondence between importance and assessability at interview. Table XLIV gives the pooled result of the 362 orders.

TABLE XLIV.—PERSONALITY AND CHARACTER TRAITS DEEMED IMPORTANT IN AIR CREW

(*Pooling of Orders in* 362 *Questionnaires*)

1. Calmness	1243
2. Dependability	1562
3. Determination	1847
4. Initiative	1920
5. Keenness	2252
6. Confidence	2290
7. Co-operation	2350
8. Discipline (= Self-discipline)	2427
9. Decisiveness	2498
10. Aggressiveness	2923
11. Respect and Influence	4082
12. Sense of Humour	4377
13. Powers of Self-Expression	4643
14. Acceptability	4692
15. Appearance and Bearing	4728
16. Breadth of Outlook	5398

This Table shows first a sharp cleavage in believed importance between traits 10 and 11, and secondly, the agreed unimportance of the traits most easily assessed in interview (e.g. traits 13, 14 and 15). The problem now takes the form: is it possible to build up a reliable standard interview in which assessments of only the traits held important are to be rated, or is it necessary to abandon the cross-table interview and substitute a much more elaborate procedure, probably on W.O.S.B. lines? It was decided to look for a solution of the first type, and a rating scale comprising traits 1–7 and 9 was drawn up. (Nos. 8 and 10 were omitted after prolonged discussion, and one fresh trait "Attitude to others," was introduced as a result of suggestions by those answering the questionnaires.) Two reliability experiments were then undertaken (1) between the assessments of interviewer and observer at a single interview, (2) between interviewers at successive interviews.

TABLE XLV.—RELIABILITY OF TRAIT ASSESSMENTS

Trait	*Independent assessments at one interview*	*Independent assessments at successive interviews*
		Interviewer A v. *Interviewer B*
	($N = 54$)	($N = 87$)
Calmness	·46	·33
Confidence	·72	·38
Dependability	·65	·28
Co-operation	·77	·60
Attitude to others	·42	·23
Decisiveness	·53	·43
Determination	·44	·42
Initiative	·58	·50
Keenness	·47	·29

The numbers in both experiments were unavoidably very small. The results on the whole conform to expectation, the coefficients in the first column being in each case higher than their counterparts in the second. The outcome of these findings is a restrained optimism about the new type of interview, on which much development work remains to be done. Meantime the following points should be noted:

(1) No interviewer undertakes the new approach without prolonged briefing and practice.

(2) No one is automatically accepted or rejected on the new interview.

(3) Apart from the introduction of a new rating scale based on a systematic analysis, the new interview differs from the old, (i) in being conducted individually, (ii) in making a more general approach (e.g. the assessment of qualities is based on a general appraisal of character rather than the candidate's attitude to Service life).

(4) It is realised that the listing of character traits may appear to savour of atomism. The importance of considering the

relationship between traits and of considering personality as a pattern is, however, strongly underlined in the briefing of interviewers.

Selection for Ground Trades

A battery of three intelligence tests (G.V.K.) and a simple arithmetic test has been given to the full ground population since 1942. The combined G.V.K. scores $\frac{(G + V + K)}{3}$ are used as a broad indicator of quality. Table XLVI is one of the guides originally given to interviewers to assist them in interpreting scores. The trades (some of them now obsolete) are divided into six groups on a "general calibre" basis. Against each group the average G.V.K. levels for "assured success" and "failure" are quoted, the columns on the right giving proportions of good (L.A.C. or A.C.1), moderate (A.C.2) and poor (ceased training) trade training results actually found for each level. The probability of success above the success level is of the order 20–1 and below the failure level only about 3–1. The scores are the averages of three percentiles based on a large random 1942 entrant population.

When the interviewer has satisfied himself on a recruit's broad grouping a study of his profile scores can often assist him in the more delicate business of recommending to a specific trade in the group. The most obvious application of profile interpretation will be the separation of High V—Low K from Low V—High K cases, the former sub-group being more likely to succeed in the clerical and the latter in the practical types of occupation.

Table XLVII gives G.V.K. multiple correlations (both before and after correction for selection) in respect of thirteen R.A.F. and eight W.A.A.F. trades. These calculations were based on training results in 1944 and 1945 the populations for each trade numbering between 100 and 300 (usually 200).

TABLE XLVI.—SHOWING AVERAGE G.V.K. SCORES FOR R.A.F. GROUND TRADES

Pre-Selection Group	*Trade*	*Average G.V.K. level for:*	*Average G.V.K. Score*	*Training results, showing % of airmen at each level who obtain these training results:*		
				L.A.C. or A.C.1	*A.C.2*	*C.T.*
(i)	Meteorologist Radio Wireless Mechanic R.D.F. Wireless Mechanic	'assured success'	80	80	20	0
		'failure'	60	0	45	55
(ii)	Photographer II Clerk Special Duties Clerk General Duties Radio Direction Finding Op. Clerk Accounting Link Trainer Instructor Wireless Op. W./T. (Slip Reader) Op.	'assured success'	70	64	34	2
		'failure'	40	12	55	33
(iii)	Flight Mechanic II Armourer (Bomb or Gun) II Instrument Repairer II Radio Telephone Op. Electrician II Armoured Car Crew Physical Training Instr. Teleprinter Op.	'assured success'	60	45	55	0
		'failure'	30	2·5	66·5	31
(iv)	Motor Boat Crew Equipment Assistant Motor Transport Mech. Motor Transport Driver Ground Observer Service Police Medical Orderly Telephone Op. Torpedoman	'assured success'	50	14	67	18
		'failure'	20	0	55	45
(v)	Ground Gunner Fabric Worker Cook and Butcher Groundsman	Failure score average G.V.K. of less than 5				
(vi)	Batman Balloon Op. Aircrafthand General Duties					

TABLE XLVII.—MULTIPLE CORRELATIONS BETWEEN G.V.K. SCORES AND TRAINING RESULTS FOR A VARIETY OF R.A.F. AND W.A.A.F. TRADES

Trade	*R.A.F. Multiple r*		*W.A.A.F. Multiple r*	
	before correction	*after correction*	*before correction*	*after correction*
Radio Wireless Mechanic	·64	·76		
Radar Operator			·33	·58
Radio Telephone Operator	·61	·68	·55	·56
Clerk Provisioning	·49	·58		
Equipment Assistant	·40	·54		
Wireless Operator	·51	—	·43	—
Wireless Telephone Slip Reader	·14	—	·37	·56
Telephone Operator	·37	·50	·36	·48
Flight Mechanic (Airframes)	·36	·47		
Flight Mechanic (Engines)	·37	·45		
Photographer	·29	·44	·46	·50
Carpenter II	·33	·40	·28	·43
Electrician II	·24	·36	·34	·52
Aircraft Finisher	·26	·32		

This Table makes clear that for certain high-grade trades G.V.K. alone can offer excellent predictive value. It was, however, never intended that it should do more than supply pointers to the trade allocation problem for which a series of more specific tests is also required. Some such tests have been used to supply corroborative evidence for special trades, but the opportunity to develop a ground trades battery analogous to the air crew aptitude battery has not yet arisen. When this does occur there will be several major problems to be tackled simultaneously with the laying-on of tests, viz.:

(1) Evaluation of the strength and meaning of trade preferences *vis-à-vis* demonstrated aptitude.
(2) Intensive briefing of interviewers as to the reliance that may be placed on the results of a more elaborate testing procedure.
(3) Separate consideration of the volunteer and National Service entrants.

The complexity of the ground trade situation (many trades, changing quotas, civilian trade experience, etc.) makes the overall validation of a selection programme inordinately difficult; but against this it may be argued that the complexity is itself a reason necessitating scientific selection. With such a wide range of human material to place and such a variety of allocations to be made the possibility of misplacement is maximal and, without the most careful planning, unavoidable.

CHAPTER XVII

CONCLUSIONS

The following series of brief statements summarises the conclusions based on personnel selection work in the Forces which we regard as having vocational or educational applications in peacetime. They are put forward somewhat dogmatically, but are accompanied by references to pages in the text where the relevant evidence or discussion has been presented.

Vocational Psychology

1.1. Psychology has many applications to contemporary human problems. It has proved itself particularly useful in the prediction of children's and adults' educational and vocational capabilities (11–12).

1.2. There is little to be learned from German vocational psychology of the 1930s, except from its mistakes. But the applications and the technical development of psychological selection methods in America are generally more extensive than in Britain. At the same time there are several aspects of American vocational psychology which should be viewed with reserve; for example, the extreme emphasis on tests, the almost universal use of selective-response test items, the analysis of abilities into independent factors, the mass production of clinical psychologists, etc. (13–24, 94–96, 169–170).

1.3. The greater advances of applied psychology in America than Britain are due partly to the greater resources of men and materials, but also to the greater prestige in which psychology is held by Government and Service authorities and by the public. Its popularity here has now much increased (probably as a result of its war-time achievements), and there are far too few qualified psychologists to meet the demands from industry, education, the Services, the universities, etc. (18, 23–24, 100–102).

Administrative and Other Considerations

2.1. The introduction of psychological methods does not involve arbitrary or bureaucratic direction of human beings. On the contrary psychologists stress the individuality of character, abilities and interests, and try to arrive at recommendations which will best suit both their subjects or candidates and the prospective employers or teachers (11–12, 85–89).

2.2. Psychologists do not wish for executive authority in carrying out selection or other schemes. They prefer to act as technical advisers, who provide the methods and train the personnel officials or teachers to apply them, but who leave the final decisions to the users. They regard it as unpsychological to impose a ready-made scheme on an industrial firm, an educational authority or an institution like the Army. They are willing to educate the users gradually, and to prove the worth of their scientific methods by the results they obtain, as they go along. At the same time they can hardly be expected to do their best work if their advice is constantly over-ruled, their methods misapplied, and if they are not allowed to administer their own affairs (27–29, 41–42, 54–55, 92–94, 100–102).

2.3. Vocational and educational psychologists have shown themselves capable of assessing social and emotional factors in their child or adult subjects, in addition to abilities, without having to refer any but seriously maladjusted cases to persons with psychiatric training. Nevertheless, so many selection, and other industrial and educational, problems involve the deeper, unconscious, factors in human motivation that co-operation between the psychological and psychiatric approaches is desirable. Another important consideration in deciding the rôle of these two professions is their relative acceptability to the users (20–21, 33, 58–60, 93–94).

Broader Psychological Aspects of Vocational Classification

3.1. Vocational or educational classification schemes should consider the needs of the institution (industrial firm, Army, etc.), and of its members as a whole. Vocational selection which merely attempts to pick the best men for one job, regardless of the needs of other branches of the institution, or of wrongfully rejected candidates, and guidance which merely recommends the most suitable jobs for single individuals, are patchwork measures which may actually increase vocational maladjustment in the long run (85-89, 92-94).

3.2. The value of vocational and educational classification schemes lies not merely in the closer co-ordination of human capacities with job or school requirements, but also in their effects on morale. Bad selection leads to lack of confidence between employers and employees, soldiers and officers, parents and schools, and possibly to neurosis. Selection that *appears* good has the opposite effects even if, judged by scientific standards, it is far from efficient. To appear good it must include consideration of each individual's interests and abilities by a sympathetic interviewer, and the tests or other instruments employed must be obviously relevant to the job (26–27, 40, 45, 52, 65–66, 97, 255).

3.3. The attitudes of the candidates towards a scheme are important, also, because their performance in psychological tests and their frankness in questionnaires or interviews are affected. It is unsafe to assume that tests which worked well in, say, the Army, would be equally effective, or that the results obtained would be duplicated, in an industrial or educational context where the motives and opinions of the candidates differ. While this applies particularly to personality factors, assessments of abilities also are not likely to be entirely immune. The confidential nature of information obtained from tests, questionnaires or interviews should be respected (65, 92–94, 100–102, 218, 257, 260).

Connection Between Classification and Training

4.1. Though classification may often have to be based on a few hours' testing and interviewing of the subjects, i.e. on a cross-section of their traits and abilities, it should preferably be integrated with training as a longitudinal process. This has the additional advantage that the training given can be adapted to the quality of the available trainees (92, 97–98).

4.2. Both classification and training are closely bound up with job simplification. The industrial psychologist, or time and motion study expert, can often so reduce the complexity of a job that a much larger proportion of candidates can manage it with simpler training (98).

4.3. Re-classification of subjects who have failed in a first job demands particularly skilful handling because of the blow to their morale. The common assumption that there is a right job for each individual, and that if he fails in one thing he possesses abilities which will make him successful at another equally complex job of a different type, may be true in some cases. More often it is necessary to recommend a lower-grade employment whose demands on the individual will be smaller (32–33, 48–49, 149).

Further Characteristics of Classification Schemes

5.1. A well-devised classification scheme does not depend on the supply of candidates falling short of, or exceeding, the number of jobs. It should also not usually involve great expenditure either of the time of highly-qualified personnel, or in the provision of elaborate tests and very detailed job analyses (85–89).

5.2. At the same time effective classification schemes cannot be run by amateur psychologists. The planning, administration and documentation, the proper choice of testing or other methods, the interpretation of their results, and in particular investigations into the worth of the methods, are inevitably highly technical matters (Chapters VI, VII, X).

Proving the Value of Classification Methods

6.1. No selection method (e.g. a test or examination), nor datum used in selection (e.g. an item from previous history, or an impression made in an interview) should be regarded as predictive of vocational or educational suitability without experimental proof. That a method works well in the "experience" of the user and seems to him to select suitable candidates shows that his attitudes to the method are favourable, but does not prove scientifically that it is a valid instrument (99–100, 119–120, 124).

6.2. It is not worth the psychologist's while to embark on a classification scheme unless he can be assured of the availability of a trustworthy and meaningful criterion of success among the people he is to classify, by comparison with which the value of his methods can be gauged. Gradings of proficiency by a single supervisor or teacher are *not* trustworthy. Written examinations at the end of a training course do *not* constitute a meaningful criterion of competence at some practical job. Nevertheless, unsatisfactory criteria can be much improved by the technical methods developed in the Forces (105–112).

6.3. It is seldom a straightforward or simple matter to demonstrate the value of a classification procedure as a whole, or of its component instruments. The main difficulties include untrustworthiness of criteria, lack of sufficient cases, disturbing influences such as variations in standards of marking or grading, and "selectivity"—i.e. the fact that it is only the candidates actually selected whose proficiency at the work can be assessed. Owing to this last factor, it is more difficult to prove the value of a procedure the more efficient that procedure is; also, any method or datum actually used in selecting is liable to appear less valid than methods not already used (107–108, 111–112, 113–117, 122–123).

6.4. Very big numbers of subjects are essential for most investigations into the value of classification methods. Several parallel studies of moderate-sized groups are preferable to a single study of one large group, because the results from different groups are liable to vary so widely (112–113).

6.5. In spite of all these difficulties the value of classification schemes has been proved beyond all doubt not only in connection with such jobs as naval, Army and A.T.S. mechanics and clerks, but also among R.A.F. aircrew. Less striking, but none the less appreciable success, was demonstrated in officer selection. There can be few jobs that call for more complex traits and abilities than these, and few, therefore, where psychological methods are unlikely to bring about improved selection (119–127, V, XVI).

The Staff Needed in Classification Schemes

7.1. The greater part of the day-to-day classification work can be effectively carried out by non-psychologists such as teachers and personnel officials. They require not merely careful selection and training but constant supervision. So great are the variations in the capacities of such persons that the value of their work should be followed up individually. Women are at least as suitable in most fields as men of equal intelligence and education, even when the psychological work is largely concerned with adult men (27–29, 41, 44–45, 73–74, 100–102, 162–163).

7.2. The training of these persons is best done by apprenticeship to a qualified psychologist. Lectures and reading can play some part, but practice in the actual application of psychological methods, and practical tests of skill before they work on their own, are essential. Work that largely involves interviewing requires if anything more training than work consisting chiefly of test administration (27–29, 41, 44–45, 73–74, 100–102).

7.3. Such persons constitute an invaluable element in a selection, or other psychological, scheme because any human affair requires tactful handling by a human being. But so great is the fallibility of human intuition and "commonsense" that they should be provided with as accurate tools as possible, and encouraged to use them in preference to their hunches. Only the exceptional human being has acumen superior to impersonal, scientifically validated, procedures (99–100, 152–164).

Provision of Information About Jobs

8.1. If candidates are supplied with accurate and easily-grasped information about available jobs, many of them are capable of a large measure of self-guidance. This not only saves the interviewer's time but provides him or her with an indication of the keenness of their interests. Interviewers also must be given information in a standard, usable form (31, 46, 90–91).

8.2. The job descriptions needed for this purpose, or as a basis for test construction, should be prepared by experts and industrial psychologists in collaboration, and should cover all the relevant physical and social conditions—not merely the operations performed. As little use as possible should be made of terms referring to generalised traits or abilities (concentration, dexterity, and the like) (47, 49, 90–91, 92–93, 167, 173).

Collection of Biographical Data

9.1. The central feature of classification procedure is the biographical questionnaire or cumulative record form where all the relevant information about previous history, test scores, interview judgments and recommendations, are brought together. Precautions in drawing up such a form are summarised on pp. 134–135 (29, 43, 91–92).

9.2. Such items as previous occupational experience, age, education, evidence of leadership and of a responsible attitude to work, and interests in particular fields, often have greater value in predicting occupational success than aptitude tests. As, however, it is difficult to elicit reliable information about these matters by written questionnaire alone, they must either be checked up in interview or, better, measured objectively by suitable tests (95–96, 135–142).

Interviewing

10.1. Conclusions relating to the conduct of the employment or diagnostic interview are summarised on pp. 143–146 (cf. also 31–32, 96–97, 146–152).

10.2. Different interviewers of the same candidates are found to arrive at widely differing conclusions, unless they are exceptionally thoroughly trained and have a clear and agreed conception as to what they are looking for. The interview judgments of psychologists and psychiatrists may have considerable value, though even these are very variable. But the average interviewer's summing up of past history and his judgments of personality qualities tend to be so subjective that they detract from, rather than add to, the accuracy of predictions based on properly weighted combinations of test scores and other objective data (152–164).

10.3. At the same time this "psychometric" approach to the assessment of vocational suitability tends to be too rigid and mechanical, and may require immense technical resources. In situations where the numbers are small, or where job requirements alter rapidly, it is quite impracticable. Its impersonality also makes it much less acceptable. The interview is, therefore, essential on the grounds of flexibility and humanity, in spite of its inaccuracy (94–97, 163–164).

Psychological Tests and Test Standards

11.1. What a psychological test appears to measure (its face value) is important from the standpoint of acceptability, but throws little light on what it actually does measure. Tests should not be regarded as eliciting hypothetical traits or abilities, but should be directly related to job proficiencies by follow-up research and their content investigated by analysis of the main common factors running through them and other tests (164–168).

11.2. Absolute standards or minimum (critical) test scores should not be laid down for acceptance for a job, or a certain type of schooling, since the state of supply and demand must be taken into account. The psychologist does not say that this man could never be successful at such and such work, but that there is a high or a low probability of his success. Follow-up information should show the optimum *ranges* of scores on the most relevant tests or groups of tests, and of other data such as age, for each job under consideration (85–89, 177–180, 278–279).

Types of Tests Most Suitable for Classification

12.1. A standard battery of group paper-and-pencil tests, containing verbal, diagrammatic or pictorial material, will by itself cover a good deal of the ground in vocational classification schemes. Tests of the information and trade knowledge type should also be available, and for certain jobs supplementary analogous or "miniature situation" tests are desirable. But our evidence indicates that it is not usually necessary to devise an elaborate battery of practical tests for measuring the aptitudes presumed to underlie each job, except in applying selection schemes to rather homogeneous samples of candidates (168–172, XII).

12.2. In constructing a new test, both the test as a whole and its component items should be shown to contribute to the prediction of vocational suitability over and above the standard battery. Economic considerations such as availability of materials and time, personnel for constructing, applying and scoring, etc., must be weighed against this contribution (172–176, 243).

12.3. Tests must be adapted to the population concerned. For example, creative-response items are preferable to selective-response in this country, and time limits should usually be generous. Tests of the abstraction and clerical type, and oral directions (provided delivery can be standardised) appear to work better among average and sub-average adults than the more conventional analogies, classification, reasoning problems, etc. Tests should not be too numerous, and most of them can and should be quite short —not more than fifteen minutes (170, 175, 220–224).

The Value of Certain Tests of Abilities

13.1. Objective spelling tests tend to be more reliable and are as valid as dictation tests of the same length, among adults. Short answer arithmetic and mathematics tests have wider predictive value than lengthy arithmetical reasoning examinations (226–229).

13.2. Such mathematics tests and verbal tests of the clerical type measure intelligence and educ*ability* in adults to a greater extent than actual education. Hence they were found to be the most useful tests of any in most Army and Navy jobs where rapid trainability and adaptability were needed (213–219, 226–229).

13.3. Mechanical comprehension, models and assembly tests, tests of spatial judgment and performance tests, tend to measure general practical intelligence in adult males rather than suitability for specifically mechanical occupations; (information tests are the most predictive for this purpose.) Such tests are more successful among adolescents and women, whose previous mechanical experience is less extensive and varied (236–246).

13.4. A combination of such mechanical and spatial tests with verbal and educational tests provides a better measure of all-round adult potentiality than do tests which aim to measure pure intelligence by non-verbal material (206, 234–236).

Factors Influencing Test Performance

14.1. Test performance is much affected by the recency of schooling and by the amount of intellectual exercise people get in their work. Thus practical abilities tend to rise during adolescence, but educational ones sink markedly except among those who receive further education, or who are engaged in intellectual jobs. General intelligence also, as measured by psychological tests, starts to decline even before the age of 20 among those who make little "use" of their "brains" (188–195).

14.2. Neither menstruation nor colds were found to have any consistent effects on the test performances of women, though evidence was obtained of improvement on certain intelligence tests among men of poor physique as a result of physical training courses (201–203).

14.3. An average rise in psychological test scores of about 5 per cent. (5 I.Q. points) is likely to occur among unsophisticated testees from having taken the same test before. Previous experience of other tests and general schooling or the taking of examinations appear to produce smaller, but appreciable, increases. Further practice produces progressively smaller effects. Hence by giving fore-exercises or one or two preliminary tests, the differences between testees with different amounts of previous experience can be much reduced. So far as the evidence goes, practice effects do not diminish with lapse of time. They are smaller in straightforward tests with ample time limits than in selective-response tests with complicated instructions and unusual content. It follows that the dangers of leakages of tests, or of previous coaching on them, are very considerable (182–187).

Assessment of Personality Qualities

15.1. War Office Selection Boards have shown the superiority of thorough study of candidates by several trained judges to ordinary interview methods of selection, and this aspect of W.O.S.B. procedure is worth applying to the selection of managers, administrative Civil Servants, and in other high-grade occupations. The inclusion of exercises, analogous to the job for which selection is taking place, helps to create confidence in the scheme, but does not necessarily improve its scientific worth. Observations of groups of candidates performing these exercises, and judgments of personality based on them, may be fully as subjective and unreliable as interview diagnoses. Far more investigation is needed into the best way of standardising the methods and making them more objective (22–23, 52–66, 122–127).

15.2. Though a definite beginning has been made in the objective measurement of important personality factors such as stability *v.* neuroticism, and extraversion *v.* introversion, most of the tests are too elaborate and time-consuming for large-scale application. Useful results are nevertheless obtainable with simple personality and interest questionnaire tests, provided that they are carefully devised in the light of the attitudes of the testees (251–254, 256–260).

15.3. Personality tests of the projection type, and indirect tests based on observation of reactions to stress, may give useful indications to trained psychologists, but are too subjective for general use. Objectively scored projection tests deserve further exploration. Among the most valuable measures of suitability for promotion to posts of responsibility and leadership are ratings or "nominations" by a man's fellow-workers (254–257).

APPENDIX I

ABBREVIATIONS EMPLOYED IN THE TEXT OR IN APPENDIX II

A.A.	Anti-aircraft (Army)
A.C.R.C.	Air Crew Reception Centre
A.C.S.B.	Aviation Candidates Selection Board
A.S.C.	Army Selection Centre
A.T.S.	Auxiliary Territorial Service (Army), now Women's Royal Army Corps
Clerks P.S.	Personnel Selection Clerks (R.A.F.)
C.O.	Commanding Officer
C.P.O.	Chief Petty Officer
C.R.C.	Combined Recruiting Centre
C.S.C.	Combined Selection Centre (R.A.F.)
C.W.	Commission and Warrant (naval officer)
D.S.P.	Directorate for Selection of Personnel (Army)
E.E.	Entry Establishment (naval recruits)
E.F.T.S.	Elementary Flying Training School
E.M.S.	Emergency Medical Service
E.S.C.	Extended Service Commission (R.A.F.)
g	General intellectual factor in ability
G.S.	General Service (Army)
G.V.K.	R.A.F. Tests of *g*, *v* and *k*
H.O.	Hostilities Only (recruits to the Navy under the Wartime Conscription Act)
I.H.R.B.	Industrial Health Research Board
I.T.W.	Initial Training Wing (R.A.F. recruits)
k	Visuo-spatial factor in ability
k : m	Spatial-mechanical factor
M.T.O.	Military Testing Officer (War Office Selection Boards)
N.I.I.P.	National Institute of Industrial Psychology
N.S.A.	National Service Act (recruits under the post-war conscription Act)
O.C.T.U.	Officer Cadet Training Unit

O.I.R.	Officer Intelligence Rating
O.T.U.	Operational Training Unit (R.A.F.)
P.S.O.	Personnel Selection Officer
P.T.C.	Primary Training Centre (army recruits)
r	Correlation coefficient
R.A.C.	Royal Armoured Corps
R.A.F.S.B.	Royal Air Force Selection Board (officers)
R.A.M.C.	Royal Army Medical Corps
R.A.S.C.	Royal Army Service Corps
R.C.A.F.	Royal Canadian Air Force
R.E.	Royal Engineers
R.E.M.E.	Royal Electrical and Mechanical Engineers
R.N.V.R.	Royal Naval Volunteer Reserve
R.T.C.	Research and Training Centre (War Office Selection Boards)
S.F.T.S.	Service Flying Training School
S.G.	Selection Group or Grade (psychological test standard)
S.P.	Senior Psychologist to the Admiralty
S.P. Test	Senior Psychologist, or Selection of Personnel, Test
T2	Total score on standard battery of naval tests
T.R.	Training Recommendation (army recruits)
U.S.A.A.F.	United States Army Air Force
v	Verbal factor in ability
v : ed	Verbal and educational factor
W.A.A.F.	Women's Auxiliary Air Force, now Women's Royal Air Force
W.O.S.B.	War Office Selection Board (officers)
W.R.N.S.	Women's Royal Naval Service

APPENDIX II

THE MAIN PSYCHOLOGICAL TESTS USED IN THE FORCES

Number or Abbreviated Title. Different series of numbers were used by Admiralty and War Office psychologists for most of their tests, but as these seldom conflicted the naval and army tests are listed together below. R.A.F. tests, to which short titles were given, appear at the end.

Use. N=used in the Royal Navy, A=the Army, a=the A.T.S., F=the R.A.F.

Time. This column gives the working time, in minutes, apart from instructions.

Reliability. These coefficients are mostly based on retesting after 1 to 6 months, and refer to representative samples of the population in the case of naval and army tests. R.A.F. samples were usually more highly selected. Italicised coefficients were obtained by the Kuder-Richardson or the split-half technique.

S.P. or D.S.P. No. or Short Title	*Title*	*Author or Source*	*Use*	*No. of Items*	*Time (mins.)*	*Reliability*
0 or 1	Progressive Matrices (1938)	Raven	Na : all candidates at C.R.C.s. A : all recruits at P.T.C.s	60	20	·88
	" " (1938)	Raven	F : volunteer recruits at C.R.C.s. Vocational Advice Service. Applicants for E.S.C.s	60	20	
	" " (1943)	Raven	F : officer candidates at R.A.F.S.B.	38	40	
1	Abstraction	Shipley (modified)	N : all recruits at E.E.s	20	10	·87
2	Mechanical Comprehension	Bennett (modified)	N : all recruits at E.E.s. Aa : all recruits at P.T.C.s	50	15	·85
3a	Arithmetic	S.P. staff	N : all recruits at E.E.s	20	2	·91
3b	Mathematics	"	" " "	30	8	
3A Pt. I	Arithmetic	D.S.P. staff.	A: all recruits at P.T.C.s	40	6	·95
3A Pt. II	Mathematics	"	" " "	38	10	
4	Squares	N.I.I.P.	N : all recruits at E.E.s. Aa : all recruits 1942-3 at P.T.C.s	50	10	·85
5	Vocabulary	Shipley	NA : experimentally, 1941-2	40	7½	
6	Comprehension, Similarities, Vocabulary	Wechsler	NA : individual tests of psychiatric cases, etc.	64	—	
7	Block Design	Kohs-Trist-Misselbrook	NA : individual tests of psychiatric cases, etc.	17	8	
8	Assembly	D.S.P. staff.	A : candidates for mechanic or driver, at P.T.C.s	9	23	·76
10	Morse Aptitude	U.S. Signal Corps	Aa : candidates for signals or operator, at P.T.C.s	78	15	·73
11	Dominoes	D.S.P. staff.	A : all recruits from 1946, at P.T.C.s	60	20	·90
12 and 21	Clerical or Instructions	D.S.P. staff.	N : occasional candidates for writers, etc. Aa : candidates for clerk, from 1943 all recruits at P.T.C.s	150	10-15	·92

S.P. or D.S.P. No. or Short Title	*Title*	*Author or Source*	*Use*	*No. of Items*	*Time (mins.)*	*Reli-ability*
14	A.T.S. Spelling	D.S.P. staff.	a : all recruits at P.T.C.s	50	5	·91
15	VIT (Verbal Intelligence Test)	D.S.P. staff.	Aa : officer candidates at W.O.S.B.s.	120	20	
16	Agility	D.S.P. staff.	A : all recruits at P.T.C.s	—	2	·58
19	Auditory Acuity	U.S.	Aa : candidates for signals, at P.T.C.s, discarded 1943	—	10	·36
22, 22A	Reading	D.S.P. staff.	A : experimental, illiterates	20		
23	A.T.S. Arithmetic	D.S.P. staff.	a : all recruits at P.T.C.s	30	10	·89
24	"Mec" Test	D.S.P. staff.	a : candidates for mechanic, at P.T.C.s	7	20	·77
25	Verbal Test	D.S.P. staff	Aa : all recruits at P.T.C.s	104	16	·97
45	Reasoning	W.O.S.B. (R.T.C.)staff	A : officer candidates at W.O.S.B.s	40	20	
70-74	Dictation	S.P. staff	N : all recruits at E.E.s	35	5	
80	Orientation	Drever (modified)	N : experimental, officer cadets	13	10	
84	Oral Test of Trade Electrical Knowledge	S.P. staff	N : occasional candidates for electrical jobs, at E.E.s	10	—	
96	Abstracting Information	U.S. Navy	N : numerous groups of officer cadets, writers, etc.	30	16	
97	Memory for Designs	N.I.I.P.	N : experimental, mechanics, artificers, etc.	8	10	
99	Radio Symbols	S.P. staff	N : experimental, candidates for radio mechanic	20		
103	Wire-bending	S.P. staff	N : experimental, mechanics	—	10	
109	Figure Construction	S.P. staff	N : experimental, mechanics	24	12	·84
110	Mechanical Models	Cox (modified)	N : experimental, artificer apprentices	7		
117E	Electrical Information	S.P. staff	N : most naval recruits from 1944, at E.E.s	25	10	·88
117M	Mechanical Information	S.P. staff	F : aircraft apprentices and boy entries	30	10	·87

S.P. or D.S.P. No. or Short Title	*Title*	*Author or Source*	*Use*	*No. of Items*	*Time (mins.)*	*Reli-ability*
118	Oscilloscope Reading	U.S. Navy	N : experimental, candidates for radar	40	10	
119	Scale and Graph Reading	S.P. staff	N : experimental, candidates for radar	48	10½	
RA	Mechanical	S.P. staff	N : all candidates at C.R.C.s, from 1946	30	15	·80
RB	Mathematics	S.P. staff	N : all candidates at C.R.C.s, from 1946	30		·94
RC	Spelling	S.P. staff	N : candidates for regular navy, at C.R.C.s	20	15	·95
RD	Abstraction	S.P. staff	N : candidates for regular navy, at C.R.C.s	20		·83
RE	Advanced Arithmetic	S.P. staff	N : candidates for branches with special educational qualifications	40	15	
RF	English	S.P. staff		30	15	
Selection Test A	Omnibus Intelligence	Vernon	N : candidates for Asdic, at Asdic Training Estabs.	75	20	
AH4	Verbal and Non-verbal Intelligence	Heim	N : candidates for Asdic	130	20	
FH3 and FHR	Verbal Intelligence and Mechanical Comprehension	Farmer, Hotoph, Rodger	A : experimental, recruits tested in units	80	20	
Group Test 25	Clerical	N.I.I.P. (abbrev.)	N : candidates for writers, etc.	141	16	
Group Test 33	Verbal Intelligence	N.I.I.P.	Na : several groups of officer candidates	193	29	
Group Test 70/1	Non-verbal Intelligence	N.I.I.P.	NA : experimental	140	3–10	
Group Test 70/23	Non-verbal Intelligence	N.I.I.P.	N : experimental, officer candidates, etc.	53	13	
Group Test 80	Spatial	N.I.I.P.	A : experimental, engineering officer candidates	80	20	

S.P. or D.S.P. No. or Short Title	Title	Author or Source	Use	No. of Items	Time (mins.)	Reliability
	Vincent Models	N.I.I.P.	F: aircraft apprentice and boy entries	8	30	
A1	Arithmetic	Stephenson.	F: all recruits at Recruit Centres; aircraft apprentice and boy entries	30	5	
G4	Non-verbal Intelligence	Stephenson.	F: a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots	59	15	·95
G5	Non-verbal Intelligence	Science 4 staff	F: volunteer recruits at C.R.C.s; recruits at Recruit Centres; aircraft apprentice and boy entries	60	20	
V4	Verbal Intelligence	Stephenson.	F: all recruits at Recruit Centres	69	15	·94
K6	Intelligence (Spatial)	Stephenson.	F: all recruits at Recruit Centres; a/crew applicants at C.S.C.; aircraft apprentice and boy entries; Navigators for Nav. (Radio)	51	15	·88
AT2	Radio Wireless Mechanic Aptitude:	Stephenson.	F: recruits for trade of Radio Wireless Mechanic			
	Pt. I—Arithmetic Problems			34	30	
	Pt. II—Technical Information			28	6	
GEN-B	Educational Attainment	Education Branch, Air Ministry	F: a/crew applicants at C.S.C.; applicants for E.S.C.s	50	20	
GEN-C	Language, Judgment and Reasoning	Science 4 staff	F: a/crew applicants at C.S.C.	65	40	
GEN-D	Aviation Information	U.S.A.A.F. (modified)	F: a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots	82	32	·89
MAT-A	General Mathematics	U.S.A.A.F.	F: a/crew applicants at C.S.C.	30	25	·82
MAT-B	Arithmetical Reasoning	U.S.A.A.F.	F: a/crew applicants at C.S.C.	30	35	·80
MAT-C	Arithmetical Approximation	U.S.A.A.F.	F: a/crew applicants at C.S.C.	15	10	·68

S.P. or D.S.P. No. or Short Title	*Title*	*Author or Source*	*Use*	*No. of Items*	*Time (mins.)*	*Reliability*
MAT-D	Speed in Simple Processes	U.S.A.A.F.	F : a/crew applicants at C.S.C.	180	10	·97
MAT-E	Calculations :					
	Pt. I—Arithmetic	Science 4 staff	F : a/crew applicants at C.S.C.	20	15	·81
	Pt. II—Mathematics		F : a/crew applicants at C.S.C.	10	12½	
MAT-F	Table Reading	U.S.A.A.F.	F : a/crew applicants at C.S.C.	90	15	·97
MEC-A	Mechanical Comprehension (pictures)	U.S.N.A.S.	F : a/crew applicants at C.S.C.	76	25	·81
MEC-B	Mechanical Comprehension (diagrams)	U.S.A.A.F. (modified)	F : a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots	41	30	·85
MEC-C	Technical Information	Stephenson	F : a/crew applicants at C.S.C.; aircraft apprentice and boy entries	29	6	·70
INS-A	Dial Reading	U.S.A.A.F. (modified)	F : a/crew applicants at C.S.C.	60	9	·90
INS-B	Instrument Comprehension	U.S.A.A.F. (modified)	F : a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots	69	32	·91
OBS-A	Aerial Photographs	U.S.A.A.F.	F : a/crew applicants at C.S.C.	48	5	·96
OBS-B	Map Reading	U.S.A.A.F.	F : a/crew applicants at C.S.C.	50	6	·84
OBS-C	Aircraft Silhouettes	U.S.A.A.F.	F : a/crew applicants at C.S.C.	60	4½	·94
MORSE-AUD	Morse Record	U.S.N.A.S. (modified)	F : a/crew applicants at C.S.C.; recruits for ground signals trades, at Recruit Centres	78	15	·62
COORD-A	S.M.A.3	Central Medical Establishment (R.A.F.) and Cambridge University	F : a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots	(apparatus)	15	·78

S.P. or D.S.P. No. or Short Title	*Title*	*Author or Source*	*Use*	*No. of Items*	*Time (mins.)*	*Reliability*
COORD-B	Control of Velocity	Cambridge University	F : a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots	(apparatus)	15	·72
COORD-C	Finger Dexterity	U.S.A.A.F. (modified)	F : a/crew applicants at C.S.C.	(apparatus)	15	·95
COORD-D	Turret Manipulation	D.Arm., Air Ministry	F : a/crew applicants at C.S.C. (experimental)	(apparatus)	15	
	Visual Link	R.C.A.F. (modified)	F : a/crew applicants at C.S.C.; applicants for Cranwell; Army Glider Pilots (experimental)	(apparatus)	3 ½-hr. periods in 1½ days	
GRADING	Flight Test :	Flying Training Command and Science 4 staffs				
	1st at 5½ hrs. dual		F : grading of pilot candidates.	Standard Flight Test		
	2nd at 8½ hrs. dual					
	3rd at 11½ hrs. dual					
NAV(R)	Navigator (Radio) :	Fighter Command and Science 4 staffs				
	(1) Problems		F : Navigators for Navigators (Radio)	10	3	
	(2) Clockface		F : ,, ,, ,, ,,	25	1	
	(3) Range and Bearing		F : ,, ,, ,, ,,	25	2	
Fitter I Maths.	Mathematics : Pt. A	Science 4 staff	F : candidates for Fitter I Conversion Course	27	12	·93
	Pt. B		F : candidates for Fitter I Conversion Course	30	15	

BIBLIOGRAPHY

ALEXANDER, W. P. (1935), "Intelligence, Concrete and Abstract," *Brit. J. Psychol. Monogr. Suppl.*, No. 29. Pp. 74.

ALLPORT, G. W. (1942), "The Use of Personal Documents in Psychological Science," *Social Sci. Res. Council Bull.*, No. 49. Pp. 210.

ANDERSON, V. V. (1929), *Psychiatry in Industry*. New York : Harper. Pp. 364.

ANSBACHER, H. L. (1941), "German Military Psychology," *Psychol. Bull.*, 38, 370–392.

BAIER, D. D. (1947), "Note on a Review of Leadership Studies with Particular Reference to Military Problems," *Psychol. Bull.*, 44, 466–467.

BALDWIN, A. L., KALHORN, J., and BREESE, F. H. (1945), "Patterns of Parent Behavior," *Psychol. Monogr.*, 58, No. 268. Pp. 75.

BALLARD, P. B. (1923), *The New Examiner*. London : University of London Press. Pp. 269.

BARR, A. S., et al. (1946), *The Measurement of Teaching Ability*. Madison, Wisc. : Dembar Publications. Pp. 206.

BENNETT, E., and SLATER, P. (1945), "Some Tests for the Discrimination of Neurotic from Normal Subjects," *Brit. J. med. Psychol.*, 20, 271–282.

BINGHAM, W. V., and FREYD, M. (1926), *Procedures in Employment Psychology*. Chicago : Shaw. Pp. 269.

BINGHAM, W. V., and MOORE, B. V. (1931), *How to Interview*. London : Harper. Pp. 320.

BION, W. R. (1946), "The Leaderless Group Project," *Bull. Menninger Clin.*, 10, 77–81.

BOARD OF EDUCATION (1924), *Report of the Consultative Committee on Psychological Tests of Educable Capacity and Their Possible Use in the Public System of Education*. London : H.M. Stationery Office. Pp. 248.

BOBBITT, J. M., and NEWMAN, S. H. (1944), "Psychological Activities at the U.S. Coast Guard Academy," *Psychol. Bull.*, 41, 568–579.

BOWERS, E. V., and WOODS, R. C. (1941), "A Study of the Experiential Factor on Intelligence Test Scores," *J. educ. Res.*, 34, 509–515.

BRITISH MEDICAL BULLETIN (various authors) (1947), "Some Problems of Personnel Research," *Brit. med. Bull.*, 5, 1, 1–57.

Brody, M. B. (1944), "Mental Testing," *Recent Advances in Psychiatry*. London : Churchill. Pp. 127–151.

Brown, J. F. (1936), *Psychology and the Social Order*. New York : McGraw Hill. Pp. 529.

Burt, C. (1921), *Mental and Scholastic Tests*. London : King. Pp. 432.

Burt, C. (1942), "Psychology in War : The Military Work of American and German Psychologists," *Occup. Psychol.*, 16, 95–110.

Burt, C. (1942–3), "The Value of Statistics in Vocational Psychology," *Occup. Psychol.*, 16, 164–174; 17, 25–33.

Burt, C. (1943a), "The Education of the Young Adolescent : The Psychological Implications of the Norwood Report," *Brit. J. educ. Psychol.*, 13, 126–140.

Burt, C. (1943b), "Validating Tests for Personnel Selection," *Brit. J. Psychol.*, 34, 1–19.

Burt, C. (1945), "The Assessment of Personality," *Brit. J. educ. Psychol.*, 15, 107–121.

Burt, C. (1947), "Symposium on the Selection of Pupils for Different Types of Secondary Schools. I. A General Survey," *Brit. J. educ. Psychol.*, 17, 57–71.

Burt, C. et al. (1926), "A Study in Vocational Guidance," *Industr. Hlth. Res. Board Rep.*, No. 33. Pp. 106.

Burt, C., and Lewis, B. (1946), "Teaching Backward Readers," *Brit. J. educ. Psychol.*, 16, 116–132.

Cantril, H. et al. (1944), *Gauging Public Opinion*. Princeton, N.J. : Princeton University Press. Pp. 318.

Cattell, R. B. (1934), "Occupational Norms of Intelligence, and the Standardisation of an Adult Intelligence Test," *Brit. J. Psychol.*, 25, 1–28.

Cattell, R. B. (1941, 1944), "An Objective Test of Character-Temperament," *J. gen. Psychol.*, 25, 59–73; *J. Soc. Psychol.*, 19, 99–144.

Cattell, R. B. (1943), "The Measurement of Adult Intelligence," *Psychol. Bull.*, 40, 153–193.

Chase, W. P. (1947), "Professional Frustration in Government Psychologists," *Amer. Psychologist*, 2, 73–75.

Chleusebairgue, A. (1939), "Industrial Psychology in Spain," *Occup. Psychol.*, 13, 33–41.

Clark, E. L. (1926), "The Value of Student Interviews," *J. Person. Res.*, 5, 204–207.

Closson, H. J., and Hildreth, H. M. (1944), "Experiment in Psychotherapy During Selection Examining," *Nav. med. Bull. Wash.*, 43, 39–43.

COCKETT, M. (1945), "A New Post-primary School Record," *Occup. Psychol.*, 19, 61–70.

COCKETT, R. (1947), "The Development of a New Test of Manual Dexterity," *Occup. Psychol.*, 21, 179–187.

CONRAD, H. S. (1947), "Psychological Tests and Their Uses : Overview and Comments," *Rev. educ. Res.*, 17, 6–9.

COX, J. W. (1928), *Mechanical Aptitude.* London : Methuen. Pp. 209.

CRAIK, K. J. W. (1945), "A New Cambridge Unit of Research in Applied Psychology," *Occup. Psychol.*, 19, 15–19.

CRITCHLEY, M. (1947), "Effects of Climatic Extremes : Critical Review," *Brit. J. industr. Med.*, 4, 164–190.

CRONBACH, L. J. (1947), "Test 'Reliability' : Its Meaning and Interpretation," *Psychometrika*, 12, 1–16.

CUNNINGHAM, K. S. (1943), "The Use of Psychological Methods in War-time Australia," *Occup. Psychol.*, 17, 111–118.

CURLE, A., and TRIST, E. L. (1947), "Transitional Communities and Social Reconstruction," *Hum. Rel.*, 1, 42–68, 240–288.

CURRAN, D., and ROBERTS, J. A. F. (1945), "A Screening Procedure for the Selection of Recruits," *J. ment. Sci.*, 91, 290–300.

DAVE, K. J. (1938), "Effects of Practice Upon Intelligence Tests," *Brit. J. educ. Psychol.*, 8, 313.

DAVIS, D. R. (1947), "Post-mortem on German Applied Psychology," *Occup. Psychol.*, 21, 105–110.

DAVIS, F. B. (1947), *Utilizing Human Talent*. Washington, D.C. : American Council on Education. Pp. 85.

DEARBORN, W. F., and ROTHNEY, J. W. M. (1941), *Predicting the Child's Development*. Cambridge, Mass. : Sci-Art. Pp. 360.

DIRECTORATE FOR SELECTION OF PERSONNEL (1947), "Personnel Selection in the British Army," *Proc. intern. Management Cong.*, 450–466.

DRAKE, C. A. (1940), "New Developments in the Selection of Factory Workers," *Selection and Development of Foremen and Workers.* New York : American Management Association.

DREVER, J., and COLLINS, M. (1936), *Performance Tests of Intelligence* (2nd ed.). Edinburgh : Oliver and Boyd. Pp. 56.

DREW, L. J. (1947), "An Investigation into the Measurement of Technical Ability," *Occup. Psychol.*, 21, 34–48.

DURANT, H., and HARRISSON, T. (1942), "Quantitative and Qualitative Method in Sociological Research," *Nature*, 149, 516–518.

DVORAK, B. J. (1947), "The New U.S.E.S. General Aptitude Test Battery," *J. appl. Psychol.*, 31, 372–376.

EARLE, F. M., and MACRAE, A. (1929), "Tests of Mechanical Ability," *Nat. Inst. industr. Psychol. Rep.* No. 3. Pp. 42.

EL KOUSSY, A. A. H. (1935), "The Visual Perception of Space," *Brit. J. Psychol. Monogr. Suppl.*, No. 20. Pp. 89.

ELLIS, A. (1946), "The Validity of Personality Questionnaires," *Psychol. Bull.*, 43, 385–440.

EXPERT COMMITTEE (1947), *Report on the Work of Psychologists and Psychiatrists in the Services.* London: H.M. Stationery Office. Pp.95

EYSENCK, H. J. (1944), "The Effect of Incentives on Neurotics, and the Variability of Neurotics as Compared with Normals," *Brit. J. med. Psychol.*, 20, 100–103.

EYSENCK, H. J. (1947a), *Dimensions of Personality.* London : Kegan Paul, Trench, Trubner. Pp. 308.

EYSENCK, H. J. (1947b), "Student Selection by Means of Psychological Tests : A Critical Survey," *Brit. J. educ. Psychol.*, 17, 20–39.

FARMER, E. (1931), "The Reliability of the Criteria Used for Assessing the Value of Vocational Tests," *Brit. J. Psychol.*, 24, 109–119.

FARMER, E., and CHAMBERS, E. G. (1936), "The Prognostic Value of Some Psychological Tests," *Industr. Hlth. Res. Board Rep.* No. 74. Pp. 41.

FARRAGO, L., et al. (1941), *German Psychological Warfare.* New York : Committee for National Morale. Pp. 133.

FEARING, F., and FEARING, F. M. (1942), "Factors in the Appraisal Interview Considered with Particular Reference to the Selection of Public Personnel," *J. Psychol.*, 14, 131–153.

FERNALD, G. G. (1912), "An Achievement Capacity Test," *J. educ. Psychol.*, 3, 331–336.

FITTS, P. M. (1946), "German Applied Psychology During World War II," *Amer. Psychologist*, 1, 151–161.

FLANAGAN, J. C. (1946), "The Experimental Evaluation of a Selection Procedure," *Educ. Psychol. Measurement*, 6, 445–466.

FLEMING, C. M. (1945), *Cumulative Records : Notes on Their Content and Uses.* London : University of London Press. Pp. 48.

FLEMING, C. M. (1948), *Adolescence.* London : Kegan Paul, Trench, Trubner.

FLUGEL, J. C. (1933), *A Hundred Years of Psychology.* London : Duckworth, Pp. 384.

FRASER, J. M. (1947), "New-type Selection Boards in Industry," *Occup. Psychol.*, 21, 170–178.

FREYD, M. (1925), "Vocational Selection," *J. appl. Psychol.*, 9, 349–356.

GARFORTH, F. I. de la P. (1945), "War Office Selection Boards, *Occup. Psychol.*, 19, 97–108.

GARRETT, H. E. (1946), "The Effects of Schooling Upon the I.Q. : A Note on Lorge's Article," *Psychol. Bull.*, 43, 72–76.

GOOD, C. V., BARR, A. S., and SCATES, D. E. (1936), *The Methodology of Educational Research*. New York : Appleton Century. Pp. 882.

GOODENOUGH, F. L., and ANDERSON, J. E. (1931), *Experimental Child Study*. New York : Appleton Century. Pp. 543.

GUILFORD, J. P. (1946), "New Standards for Test Evaluation," *Educ. Psychol. Measurement*, 6, 427–438.

GUILFORD, J. P. (1948), "Some Lessons from Aviation Psychology," *Amer. Psychologist*, 3, 3–11.

HAMLEY, H. R., et al. (1937), *The Educational Guidance of the School Child*. London : Evans. Pp. 122.

HARDTKE, E. F. (1945), "Aptitude Testing for Metalworking Occupations," *Psychol. Bull.*, 42, 679–694.

HARTOG, P., and RHODES, E. C. (1935), *An Examination of Examinations*. London : Macmillan. Pp. 81.

HEIM, A. W. (1947), "An Attempt to Test High-grade Intelligence," *Brit. J. Psychol.*, 37, 70–81.

HILL, A. B., and WILLIAMS, D. J. (1947), "Reliability of Psychiatric Opinion in the Royal Air Force," *Psychological Disorders in Flying Personnel of the Royal Air Force*. London : H.M. Stationery Office. Pp. 308–320.

HOLLIDAY, F. (1943), "The Relation Between Psychological Test Scores and Subsequent Proficiency of Apprentices in the Engineering Industry," *Occup. Psychol.*, 17, 168–185.

HOLLINGWORTH, H. L. (1929), *Vocational Psychology and Character Analysis*. New York : Appleton. Pp. 409.

HOLZINGER, K. J., and SWINEFORD, F. (1946), "The Relations of Two Bifactors to Achievement in Geometry and Other Subjects," *J. educ. Psychol.*, 37, 257–265.

HOOVERS (1948), "Selection of Supervisors : An Account of a Procedure Developed by Hoovers, Ltd.," *J. Inst. Pers. Mgmt.*, 30, 3–7.

HULL, C. L. (1928), *Aptitude Testing*. New York : World Book Co. Pp. 535.

HUNT, E. P., and SMITH, P. (1945), "Vocational Psychology and Choice of Employment," *Occup. Psychol.*, 19, 109–116.

HUNT, W. A., WITTSON, C. L., and HARRIS, H. L. (1944), "The Screen Test in Military Selection," *Psychol. Rev.*, 51, 37–46.

HUTT, M. L., and MILTON, E. O. (1947), "An Analysis of Duties Performed by Clinical Psychologists in the Army," *Amer. Psychologist*, 2, 52–56.

JAQUES, E., et al. (1947), "Social Therapy," *J. soc. Issues*, 3, No. 2.

JENKINS, W. O. (1947), "A Review of Leadership Studies with Particular Reference to Military Problems," *Psychol. Bull.*, 44, 54–79.

JENNINGS, J. R. (1947), "The Development of Personnel Selection Procedure in a Naval Entry Establishment," *Occup. Psychol.*, 21, 82–91.

JOHNSON, H. M. (1946), "The Detection and Treatment of Accident Prone Drivers," *Psychol. Bull.*, 43, 489–532.

KELLEY, T. L. (1928), *Crossroads in the Mind of Man*. Stanford, Calif. : Stanford University Press. Pp. 238.

KOHS, S. C. (1923), *Intelligence Measurement*. New York : Macmillan. Pp. 312.

KOOS, L. V. (1928), *The Questionnaire in Education*. New York : Macmillan. Pp. 178.

LINDQUIST, E. F. (1940), *Statistical Analysis in Educational Research*. Boston : Houghton Mifflin. Pp. 266.

LONG, W. F., and LAWSHE, C. H. (1947), "The Effective Use of Manipulative Tests in Industry," *Psychol. Bull.*, 44, 130–148.

LORGE, I. (1945), "Schooling Makes a Difference," *Teach. Coll. Rec.*, 46, 483–492.

LUMMIS, C. (1946), "The Relation of School Attendance to Employment Records, Army Conduct and Performance in Tests," *Brit. J. educ. Psychol.*, 16, 13–19.

McCLELLAND, W. (1942), *Selection for Secondary Education*. London : University of London Press. Pp. 264.

MACDONALD, A. (1938), "A Scheme of Vocational Guidance for Use in an Educational Area," *Occup. Psychol.*, 12, 291–301.

McNEMAR, Q. (1946), "Opinion-Attitude Methodology," *Psychol. Bull.*, 43, 289–374.

MACRAE, A. (1932), *Talents and Temperaments*. London : Nisbet. Pp. 210.

McRAE, H. (1942), "The Inconstancy of Group Test I.Q.s," *Brit. J. educ. Psychol.*, 12, 59–70.

MAGSON, E. H. (1926), "How We Judge Intelligence," *Brit. J. Psychol. Monogr. Suppl.*, No. 9. Pp. 115.

MALAMUD, D. I. (1946), "Value of the Maller Controlled Association Test as a Screening Device," *J. Psychol.*, 21, 37–43.

MANDELL, M. M., and ADKINS D. C. (1946), "The Validity of Written Tests for the Selection of Administrative Personnel," *Educ. Psychol. Measurement*, 6, 293–312.

MEIKLEJOHN, J. (1945), "Vocational Guidance in a Scottish County Area," *Occup. Psychol.*, 19, 201-211.

MEMOIRS OF THE NATIONAL ACADEMY OF SCIENCES (1921), *Psychological Examining in the United States Army*. Washington, D.C. : Government Printing Office. Vol. 15.

MERCER, E. O. (1945), "Psychological Methods of Personnel Selection in a Women's Service," *Occup. Psychol.*, 19, 180–200.

MILES, C. C., and MILES, W. R. (1932), "The Correlation of Intelligence Scores and Chronological Age from Early to Late Maturity," *Amer. J. Psychol.*, 44, 44–78.

MILES, G. H., and VINCENT, D. F. (1934), "The Institute's Tests for Motor Drivers," *Hum. Factor*, 8, 245–257.

MILLER, J. G. (1946), "Clinical Psychology in the Veterans' Administration," *Amer. Psychologist*, 1, 181–189.

MISSELBROOK, B. D. (1946), "The Short Personnel Selection Interview," *Occup. Psychol.*, 20, 85–97.

MUNROE, R. L. (1945), "Prediction of the Adjustment and Academic Performance of College Students by a Modification of the Rorschach Method," *Appl. Psychol. Monogr.*, No. 7. Pp. 104.

MYERS, C. S. (1942), "Selection of Army Personnel," *Lancet*, 243, 591–592; (1943), *Occup. Psychol.*, 17, 1–5.

NEWMAN, S. H., BOBBITT, J. M., and CAMERON, D.C. (1946), "The Reliability of the Interview Method in an Officer Candidate Evaluation Programme," *Amer. Psychologist*, 1, 103–109.

NORRIS, K. E. (1940), *The Three Rs and the Adult Worker*. Montreal : McGill University Press. Pp. 213.

NORTHCOTT, C. H. (1945), *Personnel Management*. London : Pitman. Pp. 329.

OAKLEY, C. A., et al. (1937), *Handbook of Vocational Guidance*. London : University of London Press. Pp. 337.

OLDFIELD, R. C. (1941), *The Psychology of the Interview*. London : Methuen. Pp. 144.

PALLISTER, H. (1938), "Vocational Preferences of School-leavers in a Scottish Industrial Area." *Brit. J. Psychol.*, 29, 144–166.

PARRY, J. B. (1947), "The Selection and Classification of R.A.F. Air Crew," *Occup. Psychol.*, 21, 158–169.

PATERSON, D. G., ELLIOT, R. M., et al. (1930), *Minnesota Mechanical Ability Tests*. Minneapolis : Minnesota University Press. Pp. 586.

PIOTROWSKI, Z. A., et al. (1944), "Rorschach Signs in the Selection of Outstanding Young Male Mechanical Workers," *J. Psychol.*, 18, 131–150.

POLANSKY, N. A. (1941), "How Shall a Life History be Written?" *Character and Pers.*, 9, 188–207.

POND, M. (1941), "Experience with Tests in the Scoville Manufacturing Company," *Experience With Employment Tests* (ed. H. Moore). New York : Nat. Industr. Conf. Board. Pp. 45–46.

PRICE, E. J. J. (1940), "The Nature of the Practical Factor (F)," *Brit. J. Psychol.*, 30, 341–351.

RAPAPORT, D. (1944), *Manual of Diagnostic Psychological Testing*. Publ. Josiah Macy Jr. Foundation, No. 2. Pp. 239.

RAVEN, J. C. (1939), "The R.E.C.I. Series of Perceptual Tests : An Experimental Survey," *Brit. J. med. Psychol.*, 18, 16–34.

RAVEN, J. C., and WALSHAW, J. B. (1944), "Vocabulary Tests," *Brit. J. med. Psychol.*, 20, 185–194.

REES, J. R. (1946), "Unfit on Psychiatric Grounds," *Occup. Psychol.*, 20, 125–131.

RODGER, A. (1937), "A Borstal Experiment in Vocational Guidance," *Industr. Hlth. Res. Board Rep.*, No. 78. Pp. 41.

RODGER, A. (1939a), "Planning for Vocational Guidance," *Occup. Psychol.*, 13, 1–9.

RODGER, A. (1939b), "The Use of Tests in Vocational Guidance," *Occup. Psychol.*, 13, 200–210.

RODGER, A. (1944), "The Man and His Job : From War to Peace," *Occup. Psychol.*, 18, 63–68.

RODGER, A. (1945), "The Work of the Admiralty Psychologists," *Occup. Psychol.*, 19, 132–139.

RODGER, A., NADEL, S. F., and BROWN, S. C. (1939). Chapters on Interviewing Techniques. *The Study of Society* (ed. F. C. Bartlett, E. J. Lindgren, et al.). London : Kegan Paul, Trench, Trubner. Pp. 257–271, 317–327, 379–401.

RODGER, A. G. (1936), "The Application of Six Group Intelligence Tests to the Same Children, and the Effects of Practice," *Brit. J. educ. Psychol.*, 6, 291–305.

ROGERS, C. R. (1939), *The Clinical Treatment of the Problem Child.* London : Allen & Unwin. Pp. 393.

ROLLIN, H. R. (1944), "Trade Training Failures in the W.A.A.F. : Factors in Predisposition and Precipitation," *Brit. J. med. Pyschol.*, 20, 63–76.

SARBIN, T. R. (1944), "The Logic of Prediction in Psychology," *Psychol. Rev.*, 51, 210–228.

SCOTT, I. D., and LINDLEY, C. J. (1946), "The Advisement and Guidance Program of the Veterans Administration," *Amer. Psychologist*, 1, 190–200.

SHIPLEY, W. C. (1940), "A Self-administering Scale for Measuring Intellectual Impairment and Deterioration," *J. Psychol.*, 9, 371–377.

SHUTTLEWORTH, C. W. (1942), "Tests of Technical Aptitude," *Occup. Psychol.*, 16, 175–182.

SLATER, E., and SLATER, P. (1944), "A Heuristic Theory of Neurosis," *J. Neurol. Neurosurg. Psychiat.*, 7, 49–55.

SLATER, P. (1940), "Some Group Tests of Spatial Judgment," *Occup. Psychol.*, 14, 40–55.

STANBRIDGE, R. H. (1936), "The Occupational Selection of Aircraft Apprentices of the Royal Air Force," *Lancet*, 230, 1426–1430.

STOTT, M. B. (1939), "Occupational Success," *Occup. Psychol.*, 13, 126–140.

Stott, M. B. (1943), "The Appraisal of Vocational Guidance," *Occup. Psychol.*, 17, 6–16.

Straker, D. (1944), "Difficulties of Psychological Field Research in a Fighting Service," *Occup. Psychol.*, 18, 133–141.

Strang, R. (1937), *Counseling Technics in College and Secondary School.* New York : Harper. Pp. 159.

Stuit, D. B. (ed.) (1947), *Personnel Research and Test Development in the Bureau of Naval Personnel.* Princeton, N.J. : Princeton University Press. Pp. 513.

Sutherland, J. D., and Fitzpatrick, G. A. (1945), "Some Approaches to Group Problems in the British Army," *Group Psychotherapy.* New York : American Sociometric Association. Pp. 205–217.

Symonds, P. M. (1931), *Diagnosing Personality and Conduct.* New York : Appleton Century. Pp. 602.

Taylor, H. C., and Russell, J. T. (1939), "The Relationship of Validity Co-efficients to Practical Effectiveness of Tests in Selection : Discussion and Tables," *J. appl. Psychol.*, 23, 565–578.

Thurstone, L. L. (1924), "The Principles of Vocational Guidance," *Brit. J. Psychol.*, 14, 353–361.

Thurstone, L. L. (1938), *Primary Mental Abilities.* Chicago : University of Chicago Press. Pp. 121.

Thurstone, L. L. (1944), *A Factorial Study of Perception.* Chicago : University of Chicago Press. Pp. 148.

Tiffin, J. (1946), *Industrial Psychology* (2nd ed.). New York : Prentice Hall. Pp. 386.

Traxler, A. E. (1946), "Evaluation of Aptitude and Achievement in a Guidance Program," *Educ. Psychol. Measurement*, 6, 3–16.

Trist, E. L. (1941), "Short Tests of Low-grade Intelligence," *Occup. Psychol.*, 15, 120–128.

Tuck, G. N. (1946), "The Army's Use of Psychology During the War," *Occup. Psychol.*, 20, 113–118.

Ungerson, B. (1945), "Motion Study Applied to Military Problems," Paper Read to Industrial Section, British Psychological Society, Nov., 1945.

U.S. War Department (1946), *Personnel Classification Tests.* Technical Manual 12–260. Washington : Government Printing Office.

Valentine, C. W. (1942), *Principles of Army Instruction.* Aldershot : Gale & Polden. Pp. 36.

Valentine, C. W. (1943), *The Human Factor in the Army.* Aldershot : Gale & Polden. Pp. 94.

Vernon, P. E. (1935), "Can the 'Total Personality' be Studied Objectively?" *Character and Pers.*, 4, 1–10.

Vernon, P. E. (1938a), "The Assessment of Psychological Qualities by Verbal Methods," *Industr. Hlth. Res. Board Rep.* No. 83. Pp. 124.

VERNON, P. E. (1938b), "Intelligence Test Sophistication," *Brit. J. educ. Psychol.*, 8, 237–244.

VERNON, P. E. (1939a), "Educational Abilities of Training College Students," *Brit. J. educ. Psychol.*, 9, 233–250.

VERNON, P. E. (1939b), "Questionnaires, Attitude Tests and Rating Scales," *The Study of Society* (ed. F. C. Bartlett, E. J. Lindgren, et al.). London: Kegan Paul, Trench, Trubner. Pp. 199–229.

VERNON, P. E. (1940), *The Measurement of Abilities*. London: University of London Press. Pp. 308.

VERNON, P. E. (1942), *The Training and Teaching of Adult Workers*. London: University of London Press. Pp. 48.

VERNON, P. E. (1946a), "An Experiment on the Value of the Filmstrip in the Instruction of Adults," *Brit. J. educ. Psychol.*, 16, 149–162.

VERNON, P. E. (1946b), "Statistical Methods in the Selection of Navy and Army Personnel," *J. roy. Statist. Soc.*, 8, 139–153.

VERNON, P. E. (1947a), "Research on Personnel Selection in the Royal Navy and the British Army," *Amer. Psychologist*, 2, 35–51.

VERNON, P. E. (1947b), "Psychological Tests in the Royal Navy, Army and A.T.S.," *Occup. Psychol.*, 21, 53–74.

VERNON, P. E. (1947c), "Variations of Intelligence with Occupation, Age and Locality," *Brit. J. Psychol. Statist. Sec.*, 1, 52–63.

VITELES, M. S. (1925), "The Clinical Viewpoint in Vocational Psychology," *J. appl. Psychol.*, 9, 131–138.

VITELES, M. S. (1926–30), "Psychology in Industry," *Psychol. Bull.*, 23, 631–680; 25, 309–340; 27, 567–635.

VITELES, M. S. (1932), *Industrial Psychology*. New York: Norton, Pp. 652.

WALLIN, P. (1941), "The Prediction of Individual Behaviour from Case Studies," *The Prediction of Personal Adjustment* (ed. P. Horst). *Soc. Sci. Res. Council Bull.* No. 48. Pp. 181–249.

WEBB, E. (1915), "Character and Intelligence," *Brit. J. Psychol. Monogr. Suppl.*, No. 3. Pp. 99.

WEIDER, A. MITTELMANN, B., et al. (1944), "The Cornell Selectee Index," *J. Amer. med. Assoc.*, 124, 224–228.

WILLIAMS, D. (1947), "Psychological Problems in Flying Personnel," *Brit. med. Bull.*, 5, 1, 39–43.

WILLIAMSON, E. G. (1936), "Counseling in Scholastic Motivation," *J. appl. Psychol.*, 20, 314–324.

WILLIAMSON, E. G., and BORDIN, E. S. (1940), "Evaluating Counseling by Means of a Control-Group Experiment," *Sch. and Soc.*, 52, 434–440.

WILLIAMSON, E. G., and DARLEY, J. G. (1937), *Student Personnel Work*. New York: McGraw Hill. Pp. 313.

Williamson, E. G., Wrenn, C. G., Darley, J. G., et al. (1941), "Colleges and Universities: VIII. Student Personnel Work," *Encyclopaedia for Educational Research*. New York: Macmillan. Pp. 248–326.

Wilson, A. T. M. (1946), "The Serviceman Comes Home," *Pilot Papers*, 1, No. 2, 9–29.

Wilson, N. A. B. (1945), "Interviewing Candidates for Technical Appointments or Training," *Occup. Psychol.*, 19, 167–179.

Yerkes, R. M. (1941), "Man-power and Military Effectiveness; The Case for Human Engineering," *J. consult. Psychol.*, 5, 205–209.

Yoakum, C. S., and Yerkes, R. M. (1920), *Mental Tests in the American Army*. London: Sidgwick & Jackson. Pp. 303.

INDEX

Abbreviations, 293–4
Adam, General Sir R., 40
Adjutant-General, 40
Adkins, D. C., 209, 307
Admiralty Boards, *see* Officer selection, Naval
Age, effects on proficiency, 138, 140, 158, 230, 251, 271–2, 287; *see also* Intelligence tests
Air Crew Reception Centre, 263
Air Member for Personnel, 72
Air Member for Training, 71–2
Air Ministry, 68, 71–2, 273, 299, 301
Air Training Corps, 136, 138, 140, 199–200
Alexander, W. P., 238–9, 302
Allport, G. W., 154, 302
American Army tests (1917–18), 13, 17, 206, 249
American Council on Education, 18
American selection procedures, 6, 18, 94–6, 135, 169, 213 *n.*, 214–16, 249, 257, 281; *see also* U.S.
Analysis of variance, 21, 113, 119, 136, 200, 202, 232, 284
Anderson, J. E., 134, 306
Anderson, V. V., 143, 153, 302
Ansbacher, H. L., 21, 302
Appointments Boards, 19
Army Cadet Force, 199–200
Army College, 57
Army Council, 40
Army and A.T.S. selection centres, 49, 98, 255–6
Artificer training establishment, 193
Asdic, *see* Selection of anti-submarine personnel
Atrocity stories, 50
Auxiliary Territorial Service, 39, 46–9, 53, 60, 119–20, 125, 138–9, 174, 190–1, 197–8, 202–3, 210, 212, 214, 217–18, 226–7, 229, 235–6, 239, 241–2, 248–52, 256, 285, 295–8
Aviation Candidates Selection Board, 69, 74–8, 261–2, 293

Baier, D. D., 257, 302
Baldwin, A. L., 119, 302
Ballard, P. B., 18, 302
Barr, A. S., 134, 205, 302, 306
Bartlett, Sir F. C., 68–70, 309
Bennett, E., 258, 260, 302
Bennett, G. K., 31, 43, 46 *n.*, 185, 190–1, 193–4, 201, 207, 212–14, 217–18, 230–1, 241–6, 296
Beveridge, Sir W., 39
Biesheuvel, S., 249, 254–5
Binet, A., 13, 15, 17, 18, 177, 186
Bingham, W. V., 91, 143, 172, 302
Biographical data, *see* Questionnaires
Bion, W. R., 61–2, 302
Board of Education, 14, 302
Bobbitt, J. M., 152–3, 302, 308
Bordin, E. S., 119, 311
Borstal, 118, 206
Bott, E. A., 69, 71
Bowers, E. V., 183, 302
Bowlby, J. M., 161
Boy entries (R.A.F.), 78–9, 297, 299–300
Boyd, W., 16
Boys' Brigade, 138, 199–200
Boy Scouts, 136, 138, 140, 199–200
Bracken, H. von, 154
Breese, F. H., 119, 302
British Army of the Rhine, 48, 124–5, 217
British Medical Bulletin, 38, 251, 302
British Psychological Society, 20, 54 *n.*, 101–2
Brody, M. B., 189, 303
Brown, J. F., 65, 303
Brown, S. C., 143, 309
Burt, Sir C. L., 15, 17–19, 21, 40, 50, 89, 93, 100, 116, 119, 156 *n.*, 254, 303

Cadet Forces, *see* Pre-Service organisations
Cambridge Unit of Applied Psychology, 254–5
Cambridge University Psychology Department, 6, 47, 68–70, 184 *n.*, 248, 273, 300–1
Cameron, D. C., 153, 308
Cantril, H., 152, 154, 303
Careers masters, 19–20, 87, 92, 101
Carl, G. P., 56
Carpenter, A., 184
Cattell, J. McK., 14
Cattell, R. B., 189, 195, 205, 221, 239, 255, 303

Central Medical Establishment (R.A.F.), 68, 70, 273, 300
Central Trade Test Board, 68, 70
Chambers, E. G., 19, 205–6, 305
Chase, W. P., 28, 303
Chief of Air Staff, 72
Child guidance, 13, 16, 18, 20, 93, 119, 228
Child psychology, 13, 14
Chleusebairgue, A., 175, 231, 247, 303
Church Lads' Brigade, 199
Civil Resettlement Units, 64
Civil Service, 28, 72
Civil Service Commission, 64
Civil Service Selection Board, 64
Clark, E. L., 153, 303
Class theory, 65
Clerks, Personnel Selection, 69, 73–4
Clinical psychology, 21, 281
Closson, H. J., 163 *n.*, 303
Cockett, M., 92, 304
Cockett, R., 248, 304
Collins, M., 239, 304
Combined Selection Centre, 78, 299–301
Committee of Professional Psychologists (Mental Health), 20
Commonwealth Fund, 20
Conrad, H. S., 154, 257, 304
Consultant in Neuropsychiatry (R.N.), 27
Consultant in Ophthalmology (R.N.), 27
Consultant Psychologist to Army Medical Services, 51
Continuous Service ratings, 25, 35, 239
Co-operative Test Service, 18
Correlation, multiple, 46 *n.*, 104–5, 137 *n.*, 179–80, 212–13, 216
  partial, 104, 136, 232–3
Correlation methods, 13, 15, 104, 107, 108 *n.*, 114–16, 127–30, 138, 168, 178, 213
Courtis, S. A., 15
Cox, J. W., 19, 206–7, 243, 246, 297, 304
Craik, K. J. W., 98, 248–9, 304
Critchley, M., 37, 304
Criteria of proficiency, *see* Validation
Cronbach, L. J., 105 *n.*, 304
Culpin, M., 256
Cumulative record cards, 92, 286
Cunningham, K. S., 207, 304
Curle, A., 64, 304
Curran, D., 157, 304

Darley, J. G., 20, 311–2
Dave, K. J., 183–4, 304
Davis, D. R., 23, 304
Davis, F. B., 249, 254, 304
Dearborn, W. F., 182–3, 189 *n.*, 304
Deputy Directorate of Selection (R.A.F.), 72–3 *n.*
Dickson, E. D. D., 251
Director of Armament (R.A.F.), 301
Director of Education Department (R.N.), 36
Director of Naval Recruiting, 27, 36–7
Director of Naval Training, 36
Director of Personnel Services (R.N.), 36
Directorate of Army Psychiatry, 51
Directorate of Biological Research (Army), 41, 51
Directorate for Selection of Personnel, 40–50, 53, 55–6, 119, 296–8, 304
Director-General of Army Medical Services, 41, 51
Documentation, studies of, 37, 284
Drake, C. A., 207, 304
Drever, J., 16, 40, 239, 297, 304
Drew, G. C., 70
Drew, L. J., 238, 304
Durant, H., 154, 304
Dvorak, B. J., 96, 241, 304
Dynamic psychology, 13, 16

Earle, F. M., 206, 305
Ebbinghaus, H., 15
Education, relation to proficiency, 116, 135, 138, 140–2; *see also* Tests, educational attainments
Education Branch (R.A.F.), 299
Educational psychology, 13–16, 18–21, 93, 113
El Koussy, A. A. H., 238–9, 305
Elementary Flying Training School, 262–6
Elliot, R. M., 237, 308
Ellis, A., 257, 305
Emotional instability, *see* Neurotic tendencies
Entry Establishments (R.N.), 27, 30–3, 184, 296–7
Equipment, design and layout of, 6, 13, 21, 37, 51, 98, 283
Ethical problems in applied psychology, 102, 283
Examinations, 13–14, 37, 51, 80, 87, 99, 141, 207, 228, 243, 245–6, 284
  as criteria of proficiency, 98, 106–7, 111–14, 213 *n.*
  new-type, 13, 18, 37, 81, 106, 182, 187, 213 *n.*
  reliability of, *see* Reliability of examinations
Experimental psychology, 13–14, 107, 115, 188

INDEX

Expert Committee, 27, 41, 51, 161, 305
Extravert-introvert, 252, 290
Eysenck, H. J., 171, 204 *n.*, 236, 252–4, 256, 260, 305

F factor, 238, 240
Factor analysis, 13, 15, 21, 96, 106, 111 *n.*, 167–8, 171, 209, 252, 281
Factor-content of tests, 168, 176, 221, 223, 225, 227, 234, 238, 240–1, 287
Faculties, 15, 23, 64
Failures in training, 52, 75–6, 107–8, 118–22; *see also* Transfers
Farmer, E., 19, 40, 106, 205–6, 305
Farrago, L., 21, 305
Fatigue, studies of, 13, 16
Fearing, F., 152, 305
Fearing, F. M., 305
Fernald, G. G., 253, 305
Field theory, 56, 61–2, 65, 109 *n.*
Fighter Command, 301
Films and filmstrips, for job information, 31, 47, 90
  in training, 36–7
Fitts, P. M., 122, 305
Fitzpatrick, G. A., 40, 54, 310
Flanagan, J. C., 249, 305
Fleet Air Arm, 35, 121–2, 210–11
Fleming, C. M., 92, 189, 305
Flugel, J. C., 14, 305
Flying Training Command, 301
Follow-up, 23, 34–5, 43, 64, 100–1, 109–11, 114, 117–25, 167–8, 208, 210–19, 287; *see also* Validation
Fraser, J. M., 66, 305
Freud, S., 13, 16
Freyd, M., 91, 154, 172, 302, 305

g, 15, 167–8, 171, 173, 175, 189, 191, 193–6, 199, 202, 209, 215, 221–2, 225, 227–8, 234, 236, 238–41, 243, 248, 256
Galton, Sir F., 14–15
Garforth, F. I. de la P., 53, 305
Garrett, H. E., 189, 306
Gemelli, A., 68
General Service scheme, 40, 43–6, 48, 210
German industrial and military psychology, 13, 17, 21–3, 53, 60, 64–5, 99–100, 122, 171, 208, 281
Gestalt psychology, 65, 171
Gilbreth, F. B., 16
Gillespie, R. D., 151–2, 161
Glasgow Corporation Child Guidance Clinics, 194 *n.*
Glasgow University Social and Economic Research Department, 194 *n.*
Good, C. V., 134, 306
Goodenough, F. L., 134, 306
Guilford, J. P., 105 *n.*, 254, 260, 306

H.M.S. *King Alfred*, 34, 232
H.M.S. *Royal Arthur*, 30–5, 137, 155–6, 186, 199
Hall, S., 14
Halo in ratings, 60, 99, 107, 110, 153
Hamley, H. R., 92, 306
Hardtke, E. F., 208, 306
Hargreaves, G. R., 40
Harris, H. L., 163, 306
Harrisson, T., 154, 304
Harrower, M. R., 253
Hartog, Sir P., 18, 152, 306
Harvard Growth Study, 182–3
Harvey, S. M., 256, 258
Health of Munition Workers Committee, 16
Healy, W., 16
Heard, A. O., 54 *n.*
Heim, A. W., 184, 205, 220, 231, 298, 306
Helmholtz, H. von, 14
Henmon, V. A. C., 68
Hildreth, H. M., 163 *n.*, 303
Hill, A. B., 156, 306
Holliday, F., 206, 243, 306
Hollingworth, H. L., 152, 306
Holzinger, K. J., 207, 306
Home Guard, 140
Hoovers Ltd., 66, 306
Hostilities Only ratings, 25, 27–35, 239
Hotoph, W. H. N., 120, 248, 298
Hull, C. L., 95, 169, 173–4, 306
Hunt, E. P., 87, 118, 206, 243, 306
Hunt, W. A., 163, 306
Hutt, M. L., 21, 306
Hysteric *v.* dysthymic personality, 252–4

Illiterate recruits, 131, 164, 194, 252, 256–7, 297
Individuality, psychologist's allowance for, 12, 32, 46–7, 76, 86, 97, 282
Industrial Health Research Board, 13, 16, 18–19
Industrial psychology, 14, 16, 93–4
Infantry, 125; *see also* Selection of infantry
Initial Training Wing, 263
Institute of Human Relations, *see* Tavistock Clinic
Instructional technique, *see* Training methods
Intelligence, assessment in interview, 147–8, 152
  quotients, 20, 176–7, 183, 186, 195, 289

Intelligence tests, 13, 15, 18, 22, 35, 40, 53, 56, 69–71, 79, 93, 121, 137, 139, 147, 153, 160, 166–7, 169–71, 173, 176, 180, 182–4, 187–9, 191, 204–10, 215, 220–1, 223, 233, 243, 245, 250–1, 259, 275, 288–9, 298–9
  home-made, 26, 50
  non-verbal, 17, 169, 182–4, 187, 191, 206, 215, 233–6, 238, 289, 298–9
Intelligence tests, differences between, areas of country, 188, 196–7
  groups with different interests, 136
  national groups, 196–7
  occupational groups, 136, 188, 191–2, 195–6, 205
  pre-Service organisations, 199–201
  Services, 188, 197–9
  sexes, 21
Intelligence tests, effects on, of age, 21, 176, 188–95, 221, 235, 289
  education, 189, 192–5
  health and physique, 185 *n.*, 201–3, 235, 289
  practice and coaching, 46, 174–5, 182–7, 189, 197, 201, 219, 289
Interests, assessed in interview, 147–8, 150–1
  considered in selection, 32, 43, 76, 86–7, 90–2, 95–7, 134, 140–2, 210, 215, 219, 275, 280, 282–3, 286
  tests of, 137, 171, 174, 210, 230–1, 245, 259-60, 290
Interviews, in Army and A.T.S. selection, 41, 43–7, 50, 121, 151, 156, 158, 251
  in naval selection, 25–6, 28–9, 31–4, 149–50, 155–9
  in personnel selection, 86, 91–2, 95–7, 99, 117, 134–5, 143–64, 169, 174, 208–10, 215, 252, 260, 283–7, 289–90
  in R.A.F. selection, 68, 71, 73–4, 77–80, 124, 151–2, 161–2, 276–8, 280
  in U.S. Coastguard, 153
  in U.S. Navy, 158, 163
  in W.O.S.B. selection, 52, 55–6, 58–60, 62, 126, 157, 159–61
  psychiatric, 144, 151–3; *see also* in W.O.S.B. selection

Jaques, E., 12 *n.*, 64, 306
Jenkins, W. O., 123, 205, 306
Jennings, J. R., 30, 307
Job analysis, 13, 21, 91, 112, 172–3, 207, 284, 286
  of Service employments, 36, 40, 43, 47, 49, 63, 77
Job information, 31, 36, 46–7, 49, 90–1, 149, 163, 286
Johnson, H. M., 209, 307
Junior Training Corps, 140, 199–200
Juvenile Employment Committees, 19, 118

k factor, 167, 194, 206–7, 209, 234, 238–41, 244
Kalhorn, J., 119, 302
Kelley, T. L., 238, 307
Kent, G. H., 206, 208
k : m factor, 168, 171, 194
Kohs, S. C., 32–3, 56, 194, 239–40, 246, 296, 307
Koos, L. V., 134, 307
Kuder, G. F., 259, 295

Lahy, J. M., 16
Lawshe, C. H., 95, 208, 307
Leadership, 22, 54–5, 59, 61–3, 65, 131, 138, 140, 148, 150–1, 200, 205, 209-10, 251, 286
Lewin, K., 61
Lewis, B., 119, 303
Lindgren, E. J., 309
Lindley, C. J., 20, 95, 309
Lindquist, E. F., 113, 307
Livingston, Sir P. C., 253
London County Council, 15
London School of Economics, 102
Long, W. F., 95, 208, 307
Lorge, I., 189, 307
Luftwaffe, 21
Lummis, C., 229, 307

m factor, 238
McCallum, C., 194 *n.*
McClelland, W., 108, 114, 115 *n.*, 129–30, 177, 307
Macdonald, A., 88, 307
McDougall, W., 13, 16, 254
McNemar, Q., 152, 307
MacQuarrie, T. W., 208, 237
Macrae, A., 19, 90, 100, 206, 305, 307
McRae, H., 183, 307
Magson, E. H., 152–3, 307
Malamud, D. I., 253, 307
Mandell, M. M., 209, 307
Marshall, A. J., 72
Mass Observation, 154
Mechanical ability, *see* Tests
Medical Board, Air Ministry, 68
Medical Director-General (R.N.), 36
Medical Research Council, 16, 37, 51, 184 *n.*

Meiklejohn, J., 88, 307
Memoirs of the National Academy of Sciences, 17, 307
Menstruation, effects of, 202–3, 289
Mental hospitals, psychological work in, 13, 20, 33, 225, 252–4
Mercer, E. O., 46, 307
Merit ratings, 107–11, 114–15
Merrill, M. A., 177, 186, 240
Miles, C. C., 189, 308
Miles, G. H., 231, 308
Miles, W. R., 189, 308
Military testing officers, 41, 53–5, 60–5, 126, 160
Mill Hill E.M.S. Hospital, 252–4
Miller, J. G., 21, 308
Milton, E. O., 21, 306
Ministry of Labour, 16, 19, 39, 45, 90–1, 120–1, 188
Misselbrook, B. D., 33, 56, 149, 194, 239–40, 296, 308
Mittelman, B., 311
Moore, B. V., 143, 302
Moorrees, V., 206
Morale, effects on, of selection, 27, 40, 46, 88–9, 97, 163, 282–3
  of dull recruits, 45
  of failures and transfers, 33, 35, 49, 283
  of officer candidates, 52–4, 66, 124
  of psychologists in Government employment, 28, 42, 282
  of radio mechanics, 141
  studies of, 6, 21, 24, 36, 51, 80–2
Moray House, 18, 183
Moreno, J. L., 61
Morse training, 36, 81, 137, 200, 250
Muensterberg, H., 16, 21
Munroe, R. L., 209, 308
Murray, H. A., 57
Myers, C. R., 71
Myers, C. S., 18, 40, 308

Nadel, S. F., 143, 309
Naming fallacy, 99–100, 167, 171
National Defence Research Council, 258
National Institute of Industrial Psychology, 13, 18–19, 27, 37, 41, 79 *n.*, 87, 92, 101, 117, 155, 172, 205–6, 220, 222, 231–3, 236, 296–9
National Service Act, 71, 280
Neurotic tendencies, diagnosis of, 16, 20, 31, 33, 45, 58, 94, 150–2, 155–7, 159–63, 174, 236, 239, 251–60, 290
Newman, S. H., 152–3, 302, 308
Normal frequency distribution, 15, 116 *n.*, 177 *n.*
Norris, K. E., 189, 308
Northcott, C. H., 19, 89. 308

Oakley, C.A., 90, 206, 308
Occupational Analysis Division, 96
Occupational experience, 29, 45, 47, 95–6, 131, 134–42, 147, 149–51, 210, 213, 219, 230–1, 243, 250, 286, 288
O'Connor, J., 206, 208, 253
Officer Cadet Training Units, 52, 59, 80, 122–5, 161, 257
Officer Resettlement Units, 63
Officer selection, 13, 17, 22–3, 146, 164, 174, 209, 228, 251–2, 254, 260, 285
  A.T.S., 53, 60
  Army, 39, 41, 45, 49, 52–66, 80, 109, 122–7, 134, 151, 157, 159–61, 200–1, 214, 220–1, 225, 234, 239, 255, 257, 289, 297
  naval, 25, 31–2, 34, 38, 136, 138, 159, 199, 210–14, 229, 232–3, 240, 297–8
  R.A.F., 80, 124, 296, 299–301
  U.S., 123, 153, 205, 257
Officer Training Units (R.A.F.), 263, 265
Oldfield, R. C., 146–7, 308
O'Rourke, L. J., 207
Otis, A. S., 17
Oxford University, 19, 68, 273

p factor, 241
Pallister, H., 88, 308
Parry, J. B., 75, 308
Parsons, F., 19
Paterson, D. G., 206, 237, 308
Pearson, K., 15
Percentiles, 71, 176, 180, 197, 212, 278–9
Personality, assessment by interviews, 33, 45, 74, 77–8, 96–7, 99, 143–52, 154–5, 163, 252, 260, 276–7, 285–7
  assessment by observation, 22–3, 60–2, 93, 254, 260, 290
  assessment from questionnaires, 159
  clinical and psychometric approaches, 154–5, 157–63, 285–7
  effects on tests of abilities, 215, 224, 228–9, 235
  inventories, 57, 135, 157, 210, 230–1, 252, 256–60, 290
  pointers, 56–8, 160
  ratings of, 45, 99, 109–11, 145–6, 151–3, 251, 254, 256–7, 276–7, 290

Personality—*contd.*
  structure of, 59, 69, 146, 154
  tests, 21–3, 34, 53, 148, 164, 249, 251–60, 276–8, 290
  traits, 61, 65, 91, 95, 154, 167, 169, 277–8, 286–7
Personnel selection, administration of, 28, 35–6, 41–2, 54–5, 68–9, 72–3, 281–2, 284
  in industry, 14, 16–7, 19, 29, 86–9, 92–4, 97–102, 106–8, 112–3, 115, 127, 135, 169–72, 205–10, 218–9, 260, 281–90
  matching supply and demand, 26, 32, 44, 47, 86–9, 177, 284, 287; *see also* Selection
Personnel Selection Officers, 116, 137, 149, 162, 172, 177–80, 252, 285
  Army and A.T.S., 41–50, 101, 120–2, 131, 150–1, 156–7, 200, 225, 251, 256
  naval, 27–35, 101, 121, 155–9, 199, 246
Philpott, S. J. F., 40
Physical development centres (Army), 201–2
Pilot grading, 74–8, 98, 262–72, 274–6, 301; *see also* Selection of air-crew
Pioneer Corps, 44–5
Piotrowski, Z. A., 209, 308
Polansky, N. A., 154, 308
Pond, M., 208, 308
Pre-Service organisations, 131, 136–7, 140, 199–201
Presidents of W.O.S.B.s, 41–2, 53–5, 64, 126, 160
Pressey, S. L., 258
Price, E. J. J., 238, 308
Prisoners of war, 49, 63–4, 73 *n.*
Project method, 90
Psychiatrists, in Army selection, 40, 45, 47, 49, 151, 161, 225, 239, 251, 256–7
  in child guidance, 20
  in naval selection, 31, 33, 150, 155, 157
  in personnel selection, 153, 282, 287
  in R.A.F. selection, 151, 156, 161–2
  in U.S. Coastguard selection, 153
  in W.O.S.B.s, 53, 55–60, 62, 64, 126, 159–61
  suspicions of, 33, 54, 58, 282
Psychoanalysis, 16, 33, 58–9, 73, 93–4, 282
Psychologists at W.O.S.B.s, 54–8, 109–11, 126, 159–60
Psychology, suspicions of, 11, 38, 42, 54–6, 95, 257, 281–2
Public opinion surveys, 24, 38, 144, 152, 154

Qualification form, 43, 46, 131–3, 138–9, 150
Questionnaires, biographical, 91–2, 96–7, 131–42, 144, 149, 162–3, 169, 171, 173, 210, 283, 286; *see also* Qualification form
  follow-up, 109–10
  naval, 29, 31, 35, 131, 136, 138, 140–2, 150, 159, 210
  personality, *see* Personality inventories
  R.A.F., 77
  scoring of, 95, 100, 112, 135–42, 153–5, 162–4, 174, 284, 286
  W.O.S.B., 55, 57, 134
Quetelet, L. A. J., 15

R.A.F. College, Cranwell, 80
R.A.F. Selection Boards, 80, 296
R.A.F. Selection Centres, 78–9
Rapaport, D., 20, 308
Ratings for proficiency, 48, 107, 109–11, 213, 216–7, 267–9, 284; *see also* Merit ratings; Personality
Raven, J. C., 225, 234, 236, 308–9
Reaction time, 14; *see also* Tests
Recruiting Centres, selection at, 25–7, 29–30, 40, 43, 47, 77–8, 137, 150, 155, 157, 184, 186, 197 *n.*, 198, 242, 296, 298–9
Rees, J. R., 45, 309
Rehabilitation, 49, 64, 95
Reid, D. D., 162
Reliability of examinations, 14–15, 18, 80, 98, 11, 213 *n.*
  follow-up criteria, 105–7, 111–15, 123, 125, 174, 205, 208, 217, 248, 284
  interview judgments, 152–3, 155–6, 162, 277, 290
  questionnaire items, 134–7
  tests, 47, 58, 105, 166, 176, 184, 201, 215, 226–9, 231, 235–6, 238, 243, 245, 247, 249–50, 254, 260, 270–1, 273, 288, 295–301
  W.O.S.B. procedures, 63, 122, 125–7
Research and Training Centre, 54–7, 60–1, 63–4, 122, 297
Research in vocational psychology, 19, 21, 33, 36, 50, 73, 101, 284
Rhodes, E. C., 18, 152, 306
Richardson, M. W., 295
Roberts, J. A. F., 157, 235, 304
Rodger, A., 27, 30, 40, 87–8, 90–1, 94, 102, 118, 143, 206, 298, 309

Rodger, A. G., 183, 309
Rodger, T. F., 160
Rogers, C. R., 119, 309
Rollin, H. R., 161–2, 309
Rorschach, H., 209, 253, 255–6
Rothney, J. W. M., 182–3, 189 *n.*, 304
Rowntrees Ltd., 19, 89, 208
Royal Armoured Corps, 44, 243
Royal Army Medical Corps, 40
Royal Army Service Corps, 230
Royal Artillery, 125, 139
Royal Canadian Air Force, 77,135, 301
Royal Electrical and Mechanical Engineers, 54, 57
Royal Engineers, 54, 253
Royal Marine Corps, 35, 64, 216–7, 239
Royal Naval Personnel Research Committee, 37
Russell, J. T., 130, 310

S factor, 238
Sarbin, T. R., 153–5, 309
Scates, D. E., 134, 306
Science 4, 69, 72–4, 80, 299–301
Scientific Adviser to the Air Ministry, 72
Scientific Adviser to the Army Council, 41, 51
Scientific management, 13, 16
Scott, I. D., 20, 95, 309
Scottish Council for Research in Education, 18
Sea Cadets, 138, 199–201
Seashore, C. E., 173, 209
Second Sea Lord, 27
Selection for secondary education, 18, 87, 93, 104, 115, 180, 204
  technical education, 180, 207
  university education, 18, 62–3, 171, 204, 206, 209
Selection Grades, 43, 176–7, 179–80, 185, 197–8, 212, 216–17
Selection in Allied Armies, 50
Selection of administrators, 62–3, 209–10
  African and Indian recruits, 50, 240
  air fitters, 35, 121–2, 138, 199, 210, 214
  air gunners, 69, 76, 242, 273–5
  air mechanics, 35, 121–2, 185, 199, 210, 214
  aircraft finishers, 280
  aircrafthands, 279
  aircrew, 17, 22, 68–78, 151, 164, 198–9, 210–11, 225, 229–30, 240–2, 248–9, 254, 260–78, 285, 299–301
  anti-aircraft personnel, 47, 209, 211, 217–18, 235, 239, 248

Selection of—*contd.*
  anti-submarine (Asdic) personnel, 17, 22, 25–6, 32, 34, 36, 172–3, 210–11, 224, 231–2, 247, 298
  apprentices, 35–6, 50, 78–9, 158, 206–7, 210, 236, 243, 245–6, 297, 299–300
  armoured car crews, 279
  armourers, 245, 279
  artificers, 25, 36, 193–4, 210–11, 214, 246, 297
  assembly workers, 205
  balloon operators, 279
  batmen, 279
  bombers, *see* aircrew
  builders, 44, 192, 245
  business executives, 205, 210
  carpenters, 245, 280
  cinema projectionists, 210, 246
  Civil Servants, 64, 209, 289
  clerks, 44, 70, 120, 139, 172, 179, 192, 205, 209, 213–14, 232, 259–60, 278–80, 285, 296
  coastal forces ratings, 136, 199, 217, 239
  coders, 180, 199
  communications ratings, 137–8, 199, 213–14
  compositors, 205
  cooks, 28, 139, 198–9, 210–11, 279
  cotton-mill machine fixers, 206
  deaf children, 251
  divers, 48
  draughtswomen, 239
  drivers, 16–17, 19, 44, 120, 138–9, 158, 192, 209, 214, 224, 230–1, 247, 259–60, 279, 296
  electrical mechanics, 25, 28, 36, 157–8, 210–12, 214, 228, 246
  electrical workers, 192, 245, 279–80, 297
  engineer officers, 54, 57 211–12, 236, 298
  engineers, 146, 170, 205, 243
  equipment assistants, 279–80
  fabric workers, 279
  farm workers, 192
  fitters, 207, 243, 245, 301
  flight engineers, 242, 273–6
  glider pilots, 299–301
  ground observers, 279
  groundsmen, 279
  gunnery personnel, 25, 34, 44, 175, 210–11, 213–14, 218, 247–8, 279
  infantry, 139, 213–15, 239, 243
  instructors, 32, 36, 51, 81, 210, 251, 279
  interviewers, 144, 210, 285
  labourers, 45, 192
  layers, *see* gunnery personnel

Selection of—*contd.*
  linemen, 214
  machine operators, 192, 206–7, 243
  managers, 66, 164, 257, 289
  mates, 192
  mechanicians, 36
  mechanics, 32, 44, 50, 70, 88, 120–2, 136–7, 146, 148, 171, 199, 209, 213–16, 228, 235, 239, 243–4, 260, 278, 285, 288, 296–7
    engine room, 210–11, 214
    flight, 279—80
    instrument, 214, 245, 259, 279
    motor, 210–11, 214, 245
    ordnance, 210
    see also air, electrical, radio
  metal workers, 192, 206, 208, 259
  meteorologists, 279
  motor boat crew, 279
  motormen, 205; *see also* drivers
  musicians, 209
  N.C.O.s, 32, 51, 151, 161, 211, 218, 239
  navigators, *see* aircrew
  nurses, 209
  officers, *see* Officer selection
  orderlies, 44, 279
  P.S.O.s, 41, 210
  packers, 205
  parachutists, 161
  photographers, 210–11, 279–80
  pilots, *see* aircrew
  police, 63, 152, 169, 279
  precision workers, 192
  professional workers, 66, 205, 209, 259
  psychological warfare staff, 63
  R.A.F. ground trades, 68, 70–2, 78–9, 198–9, 278–80
  radar operators, 25, 32, 34, 36, 158, 173, 175, 210–11, 230, 235, 239, 244–5, 249, 279–80, 298
  radio operators, 22, 179–80
  radio-wireless mechanics, 36, 71, 141–2, 157–8, 198–9, 210–11, 213–15, 243, 246, 279–80, 297, 299
  recruiting assistants, 27, 210
  research workers, 210
  retail tradesmen, 192
  riflemen, 44, 214
  safety equipment ratings, 138, 158–9, 210–11
  salesmen, 153, 210
  school-teachers, 205, 210
  seamen, 32, 138, 140–1, 197–9, 210–15, 228–9
  sick berth attendants, 211
  signallers and signalmen, 44, 137, 210–11, 214, 216–17, 235, 239, 244, 249–50, 296–7, 300

Selection of—*contd.*
  special operators, 120, 250
  specialists, 17, 34, 50, 62–3, 96
  stenographers, 48
  stewards, 199, 210–12
  stokers, 28, 36, 140–1, 198–9, 210–12, 214
  storemen, 44, 214
  supply assistants, 210–11, 214, 232
  switchboard operators, 139
  tank drivers, 22
  telegraphist air gunners, 35–6, 137, 210–11
  telegraphists, 17, 28, 32, 137–8, 198–9, 205, 209–11, 214, 250
  telephonists, 17, 175, 279–80
  teleprinter operators, 279
  toolmakers, 205
  torpedo ratings, 25–6, 32, 34, 36, 210–11, 218, 279
  tradesmen, 18, 25, 30, 39, 44–5, 79, 109, 120–2, 193–4, 242
  turners, 207
  W./T. slip readers, 279–80
  wireless operators, 69–70, 120, 273–4, 279–80, 296
  woodworkers, 192
  writers, 28, 32, 199, 210–12, 214, 232, 275, 296–8
Selection ratio, 86–7, 128–30, 178, 190, 216–17
Selection work abroad, 48, 50, 53, 240
Selectivity, 115–17, 122–3, 125, 137–8, 158–9, 168, 205, 212–13, 231, 244, 246, 256, 274–5, 278, 280, 284–5, 295
Senior Psychologist's Department, 27–38, 296–8
Senior Training Corps, 199–200
Service Flying Training School, 262–4, 271
Shakow, D., 206, 208
Shipley, W. C., 31, 56, 221, 225–6, 233, 296, 309
Shuttleworth, C. W., 206, 309
Slater, E., 256, 309
Slater, P., 236–7, 256, 258, 260, 302, 309
Smith, P., 87, 118, 206, 243, 306
South African Air Force, 249, 254
Spearman, C. E., 15, 222, 239
Special Entry Cadets (R.N.), 34
Special Service ratings, 25, 35
Stanbridge, R. H., 19, 309
Standardisation, *see* Tests
Stanines, 177 *n.*, 274
Statistical methods in psychology, 5, 13, 15, 38, 49, 73–4, 100, 104–5, 108 *n.*, 113, 178, 180, 284

Statistical significance, 105, 136, 192, 201–3, 274
Stenquist, J. L., 206–7, 242
Stephenson, W., 51, 68, 70–4, 78, 226, 236–8, 299–300
Stott, M. B., 106, 117, 309–10
Straker, D., 29, 107, 310
Strang, R., 143, 310
Strong, E. K., 245, 259
Student counselling, 13, 19, 89, 119
Stuit, D. B., 158, 259 *n.*, 310
Summed S.G.s, 43–4, 191, 212
Sutherland, J. D., 40, 54, 310
Sutton E.M.S. Hospital, 258
Swineford, F., 207, 306
Symonds, P. M., 134, 143, 310

Tavistock Clinic, 12 *n.*, 64, 93
Taylor, F. W., 16
Taylor, H. C., 130, 310
Temperament, *see* Personality
Tennessee Valley Authority, 207
Terman, L. M., 177, 186, 240
Tests, acceptability of, 95, 97, 163–4, 170, 189, 221, 255, 287
  analogous, 171–3
  analytic, 171–3, 209, 230–1
  apparatus, 169–70, 174, 249
  aptitude, 13, 170, 260
  compared with clinical judgments, 153–5, 157–60, 162–4, 287
  construction of, 40, 49, 73, 100, 163–4, 172–6, 286, 288
  educational attainments, 13, 15, 18, 26, 57, 74, 166, 168–9, 176, 180, 183, 189, 193–5, 199–200, 205, 213–16, 234, 238–9, 243, 245, 251, 273, 275, 288–9
  expendable, 170
  face value of, 167, 241, 255, 283, 287
  group *v.* individual, 169
  intelligence, *see* Intelligence tests
  interests, *see* Interests, tests of
  inventive *v.* selective items, 170, 175, 187, 221–2, 229, 237, 242, 281, 288
  limitations of, 94–6, 138–9, 163–4, 166–7, 218–19, 281, 287
  main characteristics of, 165–72
  mechanical scoring of, 18, 169–70, 174
  miniature situation, 172, 175, 209, 248, 288
  norms, *see* standardisation of
  origin and development of, 13–22
  paper-and-pencil *v.* practical, 47, 169–70, 207, 243–4, 248–9, 287–8
  pass-marks on, *see* standards on
  personality, *see* Personality tests
Tests—*contd.*
  practicability of, 173–4, 208, 248, 255, 259, 288
  relevance of in interviewing, 144, 147–9, 162
  reliability, *see* Reliability, tests
  screening, 157, 256; *see also* Neurotic tendencies
  standardisation of, 15, 46, 61, 65, 71, 166, 176–80, 187, 190, 195
  standards on, 29–30, 44, 49, 71, 129–30, 177–80, 266, 278, 287
  trade, *see* Trade tests
  validity, *see* Validation of tests
  weighting of, 46 *n.*, 96, 104–5, 112, 154, 174, 274
  work-sample, 62, 75, 171–2, 209, 224, 248
Tests, titles; *see also* Selection of administrators, etc.
  A1, 299
  AT2, 299
  AH4, 184, 220, 231, 298
  AH5, 205
  Abstraction, 30–1, 56, 180, 185–6, 193–5, 209, 212, 214, 217–18, 221–3, 226, 232–3, 240, 244, 246, 288, 296–8
  Agility, 43, 47, 172, 185, 190, 201, 214, 247–8, 256, 297
  Aircrew Aptitude Test Battery, 74, 76–8, 273–6
  arithmetic, *see* Tests, mathematics
  Army Alpha, 17, 195, 205, 214, 224
  Army Beta, 17, 238
  auditory, 43, 166, 173, 231–2, 250–1, 297
  Bartlett, 69, 74
  Bennett, *see* Tests, mechanical comprehension
  Bennett-Slater questionnaire, 258, 260
  Binet, 13, 15, 17, 18, 177, 186
  Body Sway, 253
  Carl Hollow Square, 56
  Cattell Scale III, 195, 205
  clerical, 43, 79, 168, 172, 178–9, 185–6, 190, 202, 212, 214–5, 217–8, 222–4, 232–3, 250, 255, 288, 296, 298
  Code Aptitude, *see* Tests, Morse Aptitude
  colour-blindness, 29
  completion, 15
  Control of Velocity, 248–9, 273, 301
  Co-ord.-A, B, C, D, 248–9, 273, 300–1
  co-ordination, *see* manual dexterity
  Cornell Selectee Index, 256, 258

Tests, titles—*contd.*
- Cox's Mechanical, 207; *see also* Tests, mechanical models
- Craik Predictor, 248
- Cube Construction, 32, 166, 205–6, 239
- Cursive Miniature Situation, 255
- dark adaptation, *see* night vision
- deterioration, 20, 221, 235
- dictation, *see* Tests, spelling
- Directions, 70, 224, 230–1, 245, 288
- Dominoes, 236, 296
- dotting machine, 254–5
- E.M.T., 69
- Echo, 249
- electrical information, 141, 158, 170, 193–4, 242–3, 246, 259, 297
- English, 30, 225, 298
- F.H.3, F.H.R., 40, 161, 220, 298
- Figure Construction, 207, 233, 237–8, 240, 297
- Fitter I Maths., 301
- flap, 249, 254–5
- Formboards, 206, 208
- Form Relations, 206, 237
- G.I.T., 69
- G.V.K., 70–1, 186, 197, 226, 236, 238, 273, 278–80, 299
- G4, 236, 273, 299
- G5, 299
- Gen. A, 273
- Gen. B, 225, 273, 299
- Gen. C, 273, 299
- Gen. D, 242, 273, 299
- General Classification, 205, 215
- general knowledge, 69, 209
- geography, 57
- graph reading, 158, 230, 244–5, 249, 298
- Group Test 25, 222, 232–3, 298
- Group Test 33, 160, 205–6, 220, 233, 298
- Group Test 70/1, 236, 244–5, 298
- Group Test 70/23, 233, 236, 244–5, 298
- Group Test 80, 207, 236, 298
- hearing, *see* Tests, auditory
- history, 57
- Ins-A, B, 230, 273, 300
- Instructions, 43, 46 *n.*, 222–4, 232–3, 236, 243, 245, 251, 255–6, 296
- K6, 238, 273, 299
- Kent-Shakow, *see* Tests, Formboards
- Kohs Block Design, 32–3, 56, 194, 239–40, 246, 296
- Kuder Preference Record, 259
- layers, 248
- leaderless groups, 34, 53, 61–2, 64–5, 126

Tests, titles—*contd.*
- Link Trainer, *see* Tests, Visual Link
- literacy, 131, 225
- Livingstone Hexagon, *see* night vision
- MacQuarrie mechanical, 208, 237
- manual dexterity, 167–8, 205, 207–9, 248, 253, 256, 273, 301
- Mat-A, B, C, D, E, F, 229, 273, 299–300
- mathematics, 30–1, 43, 46 *n.*, 48, 54, 57, 69–70, 141, 168, 179, 184–6, 189–91, 193–5, 199–201, 205–7, 212–8, 228–9, 232–4, 243–6, 249–51, 273, 278, 288, 296–301
- Matrices, *see* Tests, Progressive Matrices
- Mec (A.T.S.), 203, 242, 251, 297
- Mec-A, B, C, 242, 273, 300
- mechanical, 19, 35, 50, 88, 106, 155, 168–70, 174, 180, 184–6, 213–16, 234–5, 238–9, 298
  - assembly, 43, 57, 169, 185, 190, 201, 203, 206–7, 214, 242–3, 245, 288, 296
  - comprehension, 22, 30–1, 40, 43, 46 *n.*, 176–7, 179, 185, 190–1, 193–4, 201, 207, 212–14, 217–18, 230–1, 241–6, 273, 288, 296, 298, 300
  - information, 30, 57, 168, 170–1, 193–4, 207, 215, 242–3, 246, 273, 287–8, 297, 299–300
  - models, 193–4, 206–7, 243, 246, 288, 297, 299
- Memory for Designs, 193, 206–7, 233, 237, 244, 246, 297
- Messages, 43, 190
- Mill Hill Vocabulary, 56, 225
- Minnesota, *see* Tests, Formboards and Paper Formboard
- Moorrees, *see* Tests, Formboards
- Moray House, 18, 183
- Morse Aptitude, 43, 70, 169, 172, 176–7, 185, 190, 249–50, 273, 296, 300
- Morse Receiving, 137, 250 *n.*
- Murray, *see* Tests, Thematic Apperception
- musical talent, 173, 209
- N.D.R.C. Inventory, 258
- Nav(R), 301
- night vision, 17, 175, 209, 253, 256
- nominations, 257, 290
- numerical, *see* Tests, mathematics
- O.I.R., 56, 160
- O.R.1 Index, 46, 201
- Oakley, *see* Tests, Formboards
- Obs-A, B, C, 240–1, 273, 300

- Tests, titles—*contd.*
  - O'Connor Tweezer Dexterity, 253
  - O'Connor Wiggly Blocks, 206, 208
  - orientation, 233, 240, 297
  - oscilloscope reading, 244–5, 298
  - Paper Formboard, 207, 237–8
  - perceptual, 47, 209
  - performance, 17, 18, 22, 169, 194, 206, 238–40, 288
  - perseveration, 253
  - persistence, 253
  - personal Tempo, 253
  - physical, 168; *see also* Tests, Agility
  - physics, 57
  - Pointers, 249, 254
  - Practical Problems, 241
  - Pressey Cross-out, 258
  - Progressive Matrices, 29, 40, 43–4, 46 *n.*, 47, 56, 160–1, 168 *n.*, 176–7, 184, 186, 190–3, 195–8, 201–3, 212, 214–15, 217, 223, 230–1, 232–6, 238, 244–5, 248, 296
  - projection, 20, 22, 34, 53, 56–8, 166, 209, 255, 260, 290
  - Purdue mechanical assembly, 206
  - RA, RB, RE, RF, 30, 298
  - RC, 30, 227, 298
  - RD, 30, 221, 298
  - Radio Symbols, 246, 297
  - Raven, *see* Tests, Mill Hill; Tests, Progressive Matrices
  - reaction time, 22, 64, 175, 209, 230–2
  - reading, 171, 215, 224–5, 232–3, 297
  - Rorschach Inkblots, 209, 253, 255–6
  - S.M.A.3, 70, 248, 273, 300
  - S.P., 296–8
  - S.P. 127–30 (Dictation), 226
  - scale reading, 158, 230, 244–5, 249, 273, 298, 300
  - Seashore, *see* musical talent
  - Selection Test A, 231–2, 298
  - sensory-motor, 47, 70, 74, 171, 230–4, 254, 256
  - Shipley, *see* Tests, Abstraction; Tests, vocabulary
  - Short Format, *see* Tests, N.D.R.C. Inventory
  - shorthand, 48, 137, 172
  - sorting, 56, 109
  - spatial judgment, 167–9, 174, 183–4, 187, 206–7, 213–15, 234–40, 243, 288, 298–9; *see also* Tests, Squares
  - Speed of Response, *see* Tests, Morse Aptitude
- ests, titles—*contd.*
  - spelling, 30–1, 48, 180, 186–7, 190–1, 194–5, 214, 217–18, 224, 226–7, 229, 232, 250–1, 256, 288, 297–8
  - Squares, 31, 43, 158, 185–6, 190–1, 193–4, 207, 212–14, 217–18, 230, 236–7, 239–40, 244–6, 296
  - standard flight, 75, 78, 81, 267–9, 301
  - Stanford-Binet, *see* Tests, Binet
  - Strong Vocational Interest Blank, 245, 259–60
  - T2, 31, 35, 116, 140–2, 158–9, 180, 185–6, 191, 193–4, 197–9, 212–14, 216–18, 231–3, 244, 246, 248–9
  - tapping, 70
  - Terman-Merrill, 177, 186, 240
  - Thematic Apperception, 22, 57, 65, 166, 255
  - Track Tracer, 253–4
  - Trist-Misselbrook, *see* Tests, Kohs Block Design
  - typewriting, 48, 137
  - V4, 226, 273, 299
  - V.I.T., 56, 220–1, 233, 297
  - verbal ability, 43, 167–8, 185, 190–1, 201, 214, 224–7, 245, 273, 288, 297
  - Vincent, *see* Tests, mechanical models
  - vision, 99, 166, 174–5, 209, 256
  - Visual Link, 77–8, 301
  - vocabulary, 170, 189, 221, 224–6, 233, 296; *see also* Tests, Mill Hill; Tests, Wechsler-Bellevue
  - Wechsler-Bellevue, 33, 56, 225, 296
  - Weigl, *see* Tests, sorting
  - Wiggly Blocks, *see* Tests, O'Connor Wiggly Blocks
  - Wirebending, 242–3, 246, 297
  - Word Association, 57, 253, 255–6
- Thomson, G. H., 15
- Thorndike, E. L., 15, 68
- Thorne, General Sir A., 53
- Thurstone, L. L., 15, 112, 207, 209, 238, 310
- Tiffin, J., 86, 89, 130, 206, 310
- Time and motion study, 38, 51, 91, 283
- Toronto University, 69
- Trade tests, 18, 107, 113–14, 121, 149, 172, 246
  - oral and written, 31–2, 45, 137, 149, 169, 172, 242–3, 246, 287, 297
- Training methods, studies of, 6, 13, 15, 21, 36–7, 50–1, 71–2, 80–1, 98, 225
  - link with selection, 97–9, 283
- Training Methods department, 69, 72

Training recommendations, 43–6, 260
Training Research department, 69, 71–2, 75
Transfers to new jobs, 32, 34, 48–9, 63, 73 *n.*, 88–90, 98–9, 225, 283
Traxler, A. E., 170, 310
Treasury, 64
Trist, E. L., 33, 56, 64, 225, 239–40, 296, 304, 310
Tuck, G. N., 40, 310

U.S. Army, 123, 135, 213, 215, 249; *see also* American Army tests; American selection procedures
U.S. Army Air Force, 76 *n.*, 230, 240–2, 249, 254, 260, 273, 299–301
U.S. Coastguard, 153
U.S. Department of Labour, 96
U.S. Naval Air Service, 300–1
U.S. Navy, 158, 163, 213, 224–5, 232–3, 245, 273, 297–8
U.S. Signal Corps, 249, 296
U.S. Veterans' Administration, 20–1
U.S. War Department, 169, 178–9, 310
Ungerson, B., 51, 310
University College, London, 6

v factor, 167, 209
v : ed factor, 168, 171, 190, 194, 196, 215, 217, 221–2, 225, 227, 256
Valentine, C. W., 50, 310
Validation, 36, 86–7, 98–101, 104–30, 164
  criteria for, 81, 98, 105–17, 168, 257, 284
  of interviews, 153–5, 157–63, 284
  of questionnaire items, 135–42
  of tests, 47, 58, 75–6, 79, 95–6, 166, 172–6, 178–80, 187, 206, 208, 210, 230–3, 243–50, 254, 260, 273–4, 278, 280, 284, 287
  of W.O.S.B. selection, 55, 63–4, 66, 80
Vernon, H. M., 16
Vernon, P. E., 15, 36–7, 96, 99, 104, 107, 108 *n.*, 109 *n.*, 116, 134, 154, 166, 168 *n.*, 170–1, 182, 191, 252, 257, 298, 310–11
Vincent, D. F., 206, 231, 243, 308
Viteles, M. S., 14, 21, 91, 106, 154–5, 172, 174, 311
Vocational classification, defined, 85–9
Vocational education, 90
Vocational guidance to school-leavers, 19, 87–8, 90–2, 98, 101, 117–18
Vocational psychologists, training of, 19, 27–8, 33, 40–1, 72–4, 100–2, 277, 282, 284–5
Vocational selection and guidance, contrasted, 32, 86–7, 171–3, 208, 282
  development of, 13–23

W.E.A., 184
Wallace, J. G., 184
Wallin, P., 154–5, 311
Walshaw, J. B., 225, 309
War Office Selection Boards, *see* Officer selection
Webb, E., 152–3, 311
Wechsler, D., 33, 56, 176, 225, 296
Weider, A., 256, 311
Weigl, E., 56
Williams, D. J., 156, 162, 306, 311
Williams, G. O., 70
Williamson, E. G., 20, 119, 311–2
Wilson, A. T. M., 64, 311
Wilson, N. A. B., 91, 146–9, 159, 260, 311
Witmer, L., 16
Wittman, P., 153–4
Wittokwer, E., 160
Wittson, C. L., 163, 306
Women as vocational psychologists, 27, 29, 38, 47, 73, 285
Women's Auxiliary Air Force, 73, 161, 278, 280
Women's Royal Naval Service, 27–33, 150, 155, 157
Woods, R. C., 183, 302
Working conditions, studies of, 13, 16, 37, 47, 51, 91
Work record, *see* Occupational experience
Wrenn, C. G., 312
Wundt, W., 14

Yerkes, R. M., 11 *n.*, 17, 214, 311
Yoakum, C. S., 11 *n.*, 17, 214, 311